A Covariance Modeling Approach to Adaptive Beamforming and Detection

F.C. Robey

30 July 1991

Lincoln Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LEXINGTON, MASSACHUSETTS

91-13544

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Hugh L. Southall

Hugh L. Southall, Lt. Col., USAF
Chief, ESD Lincoln Laboratory Project Office

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

A COVARIANCE MODELING APPROACH
TO ADAPTIVE BEAMFORMING
AND DETECTION

F.C. ROBEY
Group 48

TECHNICAL REPORT 918

30 JULY 1991

LEXINGTON   MASSACHUSETTS

ABSTRACT

The subject of this report is the general problem of signal processing for sensor arrays. Under certain reasonable assumptions, the model for the noise covariance matrix of the vector of array outputs is an integral involving the spatial-temporal power-spectral-density function. This report examines the application of this covariance model to problems in adaptive beamforming and detection.

A constant false alarm rate detector, based on unconstrained maximum-likelihood techniques, is derived and analyzed. Techniques such as this do not fully exploit the data model and can show an appreciable loss in performance compared to optimal techniques.

The space of noise covariance matrices possible from a particular array is characterized, yielding representations for the space and members of the space in terms of finite numbers of spectral points. These representations are used to derive constrained maximum-likelihood estimators that jointly estimate the parameters of the density function.

Two approaches that use the constrained covariance estimates to perform beamforming are described and compared. The loss in signal-to-noise ratio and the variance of the estimators are shown to be less for these approaches than for those that do not use the covariance model.

Detection methods based on the generalized likelihood ratio test and a constant false alarm rate matched-filter detector are analyzed, and simulation results are presented. The analysis shows that the detectors will exhibit some constant false alarm rate behavior. The results show a dramatic improvement in detection performance when compared to unconstrained approaches.

Accession For

NTIS GRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	

By
Distribution/
Availability Codes

Dist	Avail and/or Special
A-1	

PREFACE

The material presented in this report is identical to that in a dissertation submitted to Washington University – Sever Institute of Technology. December 1990. in partial fulfillment of the degree of Doctor of Science.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS
(Continued)

LIST OF ILLUSTRATIONS
(Continued)

LIST OF TABLES

1. INTRODUCTION

Sensor arrays are commonly used in fields such as radar, sonar, communications, ultrasound imaging, astronomy, and geophysics. There are two basic problems which are common to these fields. The first problem is to estimate the intensity of a propagating wave from a known direction. The estimation of the intensity is a spatial filtering problem commonly known as *beamforming*. The second problem is to determine if energy is being radiated or scattered from given directions. This *detection* problem is most common in radar and sonar applications.

The interest in using sensor arrays is due to the promise of increased performance compared to that of a single-sensor system. A sensor array can provide a large physical aperture needed to produce the spatial resolution for these problems without the mechanical disadvantages of a single large antenna. A sensor array can also have the desirable feature of adjusting the spatial response of the sensor system to match a changing interference environment.

Optimal beamforming and detection methods require knowledge of the noise statistics of the environment. These statistics are seldom known and must be estimated. Beamforming and detection performance are directly dependent upon how well these estimates are made. Our interest is in improving the current beamforming and detection methods by utilizing more of the information that is available about the array and its response to the environment to form the estimates of the noise statistics.

1.1 Beamforming

Beamformers are typically designed to satisfy some optimality criteria such as maximizing the output signal-to-noise ratio, minimizing the expected value of the squared error between the estimate and the true signal, and minimizing the output variance while preserving a constant gain in the desired direction. With a Gaussian signal model, these criteria result in a linear filtering operation where the outputs of the array are weighted and summed [1]. For this reason beamforming is traditionally considered a linear filtering operation. The optimal linear beamformers for each of the criteria require knowledge of the noise covariance matrix.

1.2 Detection

Detection is required in sonar and radar applications. This is a multiple hypothesis testing problem where there could be a radiation source or scatterer in any location. Targets are often assumed to be scarce, and it is common to condition on a particular location and to perform a binary hypothesis test to determine the presence or absence of a target at that location. When the noise statistics are known and assumed to be Gaussian, this results in the comparison of the output power of an optimal linear beamformer to a fixed threshold. When the noise statistics are unknown, the parameters of the beamformer are unknown, and adaptive approaches to detection must be utilized.

1

Several quantities are used to quantify detector performance. The first of these is the probability of false alarm (PFA), which is the probability that the presence of a signal is indicated when it is not present. The second performance measure is the probability of detection (PD). This is the probability that the detector indicates a signal is present when it is. For a particular detector these probabilities are related: typically, lowering the false alarm probability also reduces the probability of detection at a particular signal level.

We are typically interested in optimizing detection performance based on the Neyman-Pearson criterion [2,3]. The Neyman-Pearson criterion is to maximize the probability of detection while restricting the probability of false alarm to be less than or equal to a fixed value. A likelihood ratio test, when it exists, satisfies this criterion.

A desirable property of detectors is independence of the probability of false alarm with respect to any unknown noise parameters. A detector which exhibits this property is considered to be a constant false alarm rate (CFAR) detector.

1.3 Parameter Estimation

The parameters of the noise environment must be known in order to perform optimal beamforming and detection. In general the statistics of the noise environment are not known and change with time. If the weight vector used to form the linear filter is fixed, then there can be a significant loss in signal-to-noise ratio compared to the optimal weight vector for the noise environment. The probability of false alarm for a detector is extremely sensitive to the noise covariance matrix, and a relatively small increase in the noise level can result in a large increase in the PFA. For this reason beamformers and detectors must be able to adapt to the changing noise environment. The coefficients of the beamformer or the parameters of the detector must be estimated from the data that are available.

It should be emphasized that the only reason we would want to estimate the noise parameters is that they affect the performance of the beamformers and detectors. The noise parameters are nuisance parameters.

There are two basic methods of adapting to the changing noise environment. The first method is to initialize the beamformer and then estimate the weight vector as data are received. This results in the weight vector being recursively updated for each input data point. The second method of dealing with the changing noise environment is to utilize all of the data available up to that point to derive the test or beamformer. I will refer to methods that do this as block adaptive as they operate on the entire block of available data at one time. The emphasis of this report is on block adaptive approaches.

1.4 Outline of This Report

The first three chapters of this report are devoted to introductory material concerning the data model and current methods of performing beamforming and detection with sensor arrays. A

2

detection method that is based on current techniques is the subject of Chapter 4. In Chapters 5 through 7, constrained covariances and estimation methods are discussed. The use of constrained covariance estimates to perform beamforming and detection and a comparison of the performance for these methods are discussed in Chapters 8 and 9.

The data model is introduced in Chapter 2. This model is based on the interference environment and the processing which occurs in a matched-filter receiver. The temporal samples of the receiver outputs to a deterministic signal are shown to be array "direction vectors." A stochastic model for the interference is introduced which results in an integral expression for the covariance matrix, involving the spatial-temporal power-spectral density. This expression can be used to show that for some arrays the covariance matrix will exhibit structure, and an example is given. The goals of the beamformers and the detectors that are proposed in this report are discussed.

Current methods of beamforming and detection are discussed in Chapter 3. The beamformers discussed here have been derived with the intent of optimizing some aspect of the beamformer performance. Adaptive beamformers are discussed. The adaptive detection methods that are discussed in this chapter include the "cell-averaging" approaches and approaches based on generalized likelihood ratio techniques.

An adaptive detection method based on the known covariance colored-noise matched-filter detector is described in Chapter 4. This method utilizes the unconstrained maximum-likelihood estimate of the noise covariance matrix and provides a CFAR test. The test statistic for this approach is simpler than the generalized likelihood ratio test statistic and exhibits similar detection performance.

In Chapter 5 the covariance structure introduced by the integral expression for the covariance matrix is investigated. This provides methods of representing the space of possible covariance matrices and individual covariance matrices that are members of that space. These representations can be used in practical implementations of covariance estimators.

An estimator that restricts the possible space of covariance matrices to those possible with a given array is developed in Chapter 6. This estimator jointly estimates the mean and the covariance of the Gaussian density function. The resulting iterative estimator has the property that the likelihood for each iteration is a non-decreasing sequence, and estimates that satisfy the necessary conditions for the maximizer of the likelihood are stable points.

In Chapter 7 a second estimator that jointly estimates the mean and structured covariance matrices is proposed. This estimator provides computational advantages compared to the estimator of the previous chapter. This estimator can require less computation per iteration, and the convergence of the estimates is more rapid for some interference environments.

Methods of utilizing the estimation procedures of the previous chapters to perform beamforming and detection are proposed in Chapter 8. Two beamforming and two detection methods are proposed and discussed. The detection methods discussed in this chapter have some CFAR

properties and result in tests where the performance is much closer to that which is provided by the known noise statistics test as well.

In Chapter 9 the beamforming and detection methods are compared by use of computer simulations. It is shown that a dramatic improvement in performance can be obtained by using the procedures proposed in this report. A summary of the results is discussed in Chapter 10.

2. DATA MODEL AND PROBLEM STATEMENT

In this chapter, the data model will be derived, and formal problem statements will be made. The data model is based on the physical and electrical characteristics of the operating environment and is similar in both the radar detection problem and the communications problem. The data model for the radar problem will be derived in detail, and the modifications needed for the communications data model will be indicated.

2.1 Sensor Array in Noise Field

It is assumed that there is an array of sensors arbitrarily spatially located in an environment with propagating electromagnetic waves. The complex envelope of a propagating wave whose properties are to be examined is considered a signal, and all other propagating waves are considered noise. The sensors will respond to the propagating wavefield yielding spatial samples of the field. This is shown pictorially in Figure 1. This report is not concerned with different propagation modes
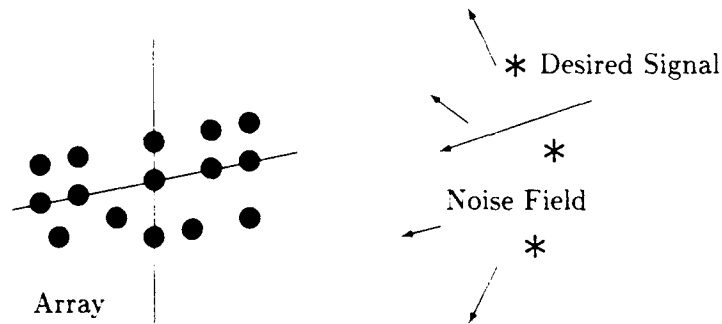
Figure 1. A sensor array with incident signals and noise.

of the waves; therefore, it is assumed that the sensors respond equally to different polarizations of the propagating waves. The response of the receivers with a deterministic signal will be discussed initially, and then the model that results when noise is included will be examined.

2.2 Deterministic Signal

It will be assumed that a pulsed signal is emitted by a radar transmitter, reflected or scattered by a single point target, and then sensed by a receiver. The propagation medium is assumed to be homogeneous, and no other reflectors or targets are present. The signal transmitted is

5

$$s(t) = \begin{cases} \sqrt{E_t}\,\Re\{A(t)e^{j\omega_o t}\} & -T/2 \le t \le T/2 \\ 0 & \text{otherwise} \end{cases} \quad . \tag{1}$$

$A(t)$ is the complex modulation of the transmitted signal and is assumed to be normalized so that

$$\int_{-T/2}^{T/2} |\Re\{A(t)e^{j\omega_o t}\}|^2 dt = 1 \quad . \tag{2}$$

$A(t)$ is typically used to create a transmitted signal with a large time bandwidth product. ω_o is the carrier or center frequency, and E_t is the transmitted energy.

The complex envelope of the transmitted signal will experience the combination of the following effects before being received by sensor i.

1. Signal attenuation G_p and propagation delay $\tau_r(t)$ and $\tau_i(\bar{e})$. The signal attenuation is due to path loss and combines the loss for propagation from the array to the scatterer and then again from the scatterer back to the sensor. G_p is typically proportional to $\frac{1}{r^2}$ corresponding to a spherical wavefront for both propagation directions. The propagation delay is due to the non-zero time that it takes for the wave to propagate from the array to the scatterer and back. The scatterer is assumed to be at a range $R(t) = R_o + R' t$ for range R_o and relative velocity R'. The delay from the array reference to the target and back to the array reference will be approximated as $\tau_r(t) = \tau_o + 2R' t/c$ for target relative velocity R' much less than the propagation velocity c and where $\tau_o = 2R_o/c$. There is an additional differential delay term $\tau_i(\bar{e})$ due to the translation of the sensor away from the array reference. This delay $\tau_i(\bar{e})$ is dependent upon the sensor location $\mathbf{s}_i$ and the direction of propagation $\bar{e}$. For heterogeneous media, there would typically be multipath and refraction effects which would need to be taken into account here.

2. Target reflectivity $\rho e^{j\vartheta}$. It is assumed that the target scatters the signal and imparts a phase shift and scale factor to the complex envelope of the incident signal.

3. Sensor response $B_i(\bar{e})$. The sensor response is a function of the spatial origination of a propagating wave. It will be assumed that the target is in the sensor far field so that the sensor response is a function only of the direction of propagation. The sensor is assumed to have a bandwidth that is much wider than that of the receiver and the transmitted signal; the gain is then independent of the temporal frequency of the received waveform.

Under the model described above, the signal received by sensor i can be written

$$x_i(t) = \Re\{\sqrt{E_t}G_p\rho e^{j\vartheta}B_i(\bar{e})A(t - \tau_o - 2R' t/c - \tau_i(\bar{e}))e^{j\omega_o(t - \tau_o - 2R' t/c - \tau_i(\bar{e}))}\} \quad . \tag{3}$$

6

A received signal amplitude term b_i can be defined combining the effects of path loss G_p, target reflectivity $\rho e^{j\vartheta}$, the transmitted energy $\sqrt{E_t}$ and part of the propagation induced phase shift $e^{j\omega_o T_o}$. This received signal amplitude is

$$b_i \equiv \sqrt{E_t} G_p \rho e^{j\vartheta} e^{-j\omega_o T_o} \tag{4}$$

and is a complex parameter.

It will be assumed that all of the receivers are identical matched-filter receivers and that the outputs of the receivers are time samples of the in-phase (I) and quadrature (Q) components of the received signal. The matched filtering in the receiver and the modulation $A(t)$ are designed so that the received signal will be time compressed before it is sampled with the result that the signal will appear in only one time sample.[1] This time compression is used to increase the signal-energy-to-interference-energy ratio in one of the time samples. A block diagram of a receiver which can accomplish this is shown in Figure 2. The combined effect of the radio frequency (RF),
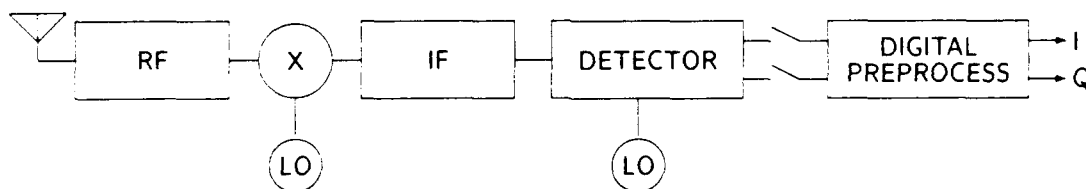
Figure 2. Simplified receiver.

---

[1]For the purposes of this report, the actual location of the compression and the implementation are unimportant. The matched filtering that implements the compression may be located in the RF, IF, video stages or digitally after sampling the video. In practice, the actual location of the compression may be important due to dynamic range and other implementation considerations.

intermediate frequency (IF), and the detector sections is to implement the matched-filter receiver. The time compression of the complex envelope due to target motion will be assumed to have a negligible effect in the matched filtering.

A block diagram illustrating the equivalent processing accomplished for a single receiver is shown in Figure 3. All operations that are performed after the multiplication by the complex
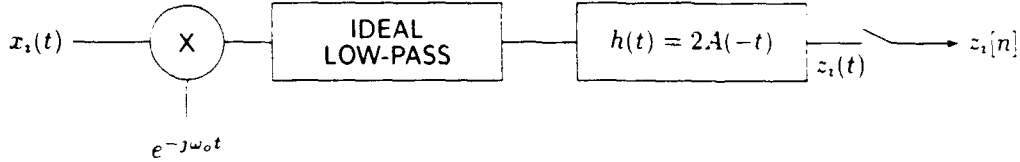
Figure 3.   Receiver signal processing.

exponential are operations on the resulting complex waveform. The signal $z_i[n]$ will be

$$z_i[n] = \left\{ [x_i(t)\, e^{-j\omega_o t}]_{LP} ** h(t) \right\}_{nT_s} \qquad , \qquad (5)$$

where $[\cdot]_{LP}$ stands for an ideal low-pass function, and $**$ is the convolution operator. The sampling period is T_s. For a matched-filter receiver the filtering in the receiver has an impulse response which is the time reversed transmitted waveform $h(t) = 2A(-t)$. Substituting $x_i(t)$ and $h(t)$ into this expression yields the result that

$$z_i[nT_s] = \begin{cases} b_i B_i(\bar{\mathbf{e}}) e^{-j\omega_o \tau_i(\bar{\mathbf{e}})} e^{-jnT_s\omega_o 2R'/c} & nT_s = \tau_o \\ 0 & nT_s \neq \tau_o \end{cases} \qquad , \qquad (6)$$

8

where the change in phase $-\omega_o 2R' t/c$ has been assumed to be negligible over the period of the modulation. $A(t)$ is assumed to have been chosen so that $z_i[nT_s] = 0$ for $nT_s \neq \tau_o$.

If transmitted pulses are spaced T_p seconds apart, then the received data for range sample n of pulse m are

$$z_i[nT_s] = \begin{cases} b_i B_i(\bar{e}) e^{-j\omega_o \tau_i(\bar{e})} e^{-jnT_s \omega_o 2R'/c} e^{-j\omega_o m T_p 2R'/c} & nT_s = \tau_i \\ 0 & nT_s \neq \tau_i \end{cases} \qquad (7)$$

Unless otherwise noted, it will be assumed that there is a single pulse: $m = 0$. A redefinition of b_i can now be made to incorporate the phase shift $-nT_s \omega_o 2R'/c$.

$$b_i \equiv \sqrt{E_t} G_p \rho e^{j\vartheta} e^{-j\omega_o \tau_o} e^{-jnT_s \omega_o 2R'/c} \qquad . \qquad (8)$$

When there are multiple samples, the temporal samples corresponding to the same delay or range can be ordered, and these samples can be further filtered into several Doppler "bins." The methods presented in this report can be applied either to the outputs of the Doppler bins or to the temporal samples directly.

2.3 Deterministic Signal for Multiple Receivers

With the output of a single receiver defined, the output of the receivers for N spatially separated sensors can now be investigated. It will be assumed that the target is located in the array far field so that the signal wavefront can be approximated by a plane wave.

If the magnitude of the complex envelope of the received signal does not vary over the array aperture, then the signal is described as *narrowband*. The signal will experience a time delay with respect to the array reference dependent upon the relative position and orientation of the sensor, the direction of arrival of the wave, and the speed of propagation. In Euclidean coordinates, let $\bar{e}$ be a unit vector in the direction of propagation of the wave, and let s_i be a vector from the array reference to sensor i. For the signal received at sensor i, the resulting delay with respect to the array reference will be $-(\bar{e} \cdot s_i)/c$ where "$\cdot$" indicates the standard Euclidean dot-product, and c is the propagation velocity. Figure 4 illustrates this geometrically for a plane containing $\bar{e}$ and s_i. Also shown in this figure is the wavelength, $\lambda_o = 2\pi c/\omega_o$, of the traveling wave. It will be convenient to express distances in wavelengths later in this report.

The data samples from multiple receivers at time n can be arranged in a vector. The vector output of the array due to the received signal will then be a complex scalar times an array steering or direction vector:

$$\mathbf{z}[n] = b\mathbf{d}(\omega, \bar{e}) \qquad . \qquad (9)$$

9
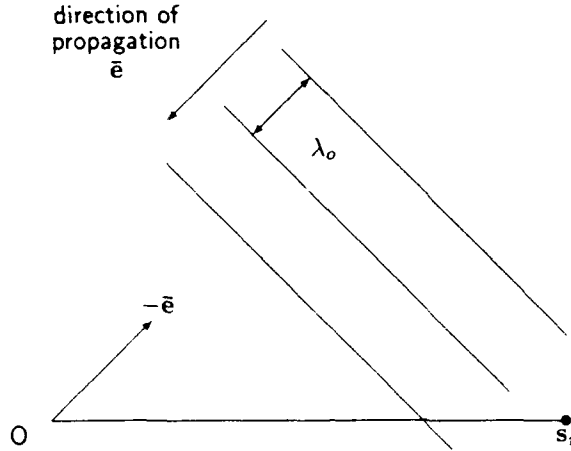
Figure 4.   Plane wave propagating through array.

The array steering vector $\mathbf{d}(\omega, \bar{\mathbf{e}})$ is

$$\mathbf{d}(\omega, \bar{\mathbf{e}}) \equiv [d_1 \; d_2 \; \cdots \; d_N]^T \qquad (10)$$

with elements

$$d_i = B_i(\bar{\mathbf{e}})e^{-j\omega_o \tau_i(\bar{\mathbf{e}})} = B_i(\bar{\mathbf{e}})e^{j\frac{\omega_o}{c}\bar{\mathbf{e}} \cdot \mathbf{s}_i} \qquad . \qquad (11)$$

Unless necessary to prevent misunderstanding, the dependence of the direction vector $\mathbf{d}(\omega, \bar{\mathbf{e}})$ on ω and $\bar{\mathbf{e}}$ will be suppressed.

2.4   Stochastic Sources

In addition to the deterministic signal which may be present in the environment, there will also be additive noise sources present. These additive noise sources can be divided into two categories. The first category is noise that is generated within the receiver due to thermal and other effects. The sum of the additive noise terms that are generated within the receiver will be denoted $\mathbf{n}_r[t]$.

10

The second noise category is noise that originates in the environment and propagates as waves to be sensed along with any signals that may be present. Natural background noise is always present at some low level due to cosmic radiation, sky noise, solar radiation, and blackbody radiation of the earth. There may also be interference due to other radio emissions. The contribution of the spatially distributed noise sources to the receiver outputs will be denoted $\mathbf{n}_s[t]$.

The following assumptions will be made.

1. The spatially generated noise originates from a zero mean stochastic process indexed by t and $\bar{\mathbf{e}}$. Anticipating the bandpass filtering in the receiver, this process is assumed to be a low-pass process modulating a sinusoid at the carrier frequency and can be written $\tilde{n}(t, \bar{\mathbf{e}})$.

2. The noise process is assumed to be temporally wide sense stationary and spatially white:

$$\mathrm{E}\{\tilde{n}(t_1, \bar{\mathbf{e}}_1)\tilde{n}^*(t_2, \bar{\mathbf{e}}_2)\} = \tilde{r}(t_1 - t_2, \bar{\mathbf{e}}_1)\delta(\bar{\mathbf{e}}_1 - \bar{\mathbf{e}}_2) \qquad . \qquad (12)$$

The spatial-temporal power-spectral density is

$$\tilde{S}(\omega, \bar{\mathbf{e}}) = \int_{\infty}^{\infty} \tilde{r}(\tau, \bar{\mathbf{e}})e^{-j\omega\tau}\,d\tau \qquad . \qquad (13)$$

3. It is assumed that the sources of interference are located in the array far field. This assumption is reasonable since systems designers typically attempt to minimize the interference to a receiver system.

Under these assumptions. the spatially generated noise that is sensed at a receiver may be written as the integral

$$x_i(t) = \Re\left\{\int_{\bar{\mathbf{e}}\in\Omega_{\bar{\mathbf{e}}}} B_j(\bar{\mathbf{e}})\tilde{n}(t - \tau_i(\bar{\mathbf{e}}), \bar{\mathbf{e}})e^{j\omega_o(t-\tau_i(\bar{\mathbf{e}}))}\,d\bar{\mathbf{e}}\right\} \qquad . \qquad (14)$$

This integral is over the set of all possible directions of arrival of the interference $\Omega_{\bar{\mathbf{e}}}$. τ_i here is with respect to the array reference and by the far-field assumption $\tau_i(\bar{\mathbf{e}}) = -\frac{1}{c}\bar{\mathbf{e}} \cdot \mathbf{s}_i$.

Van Trees [4] has derived a complex representation for bandpass random processes for a single receiver. This representation and the statistics can be utilized for multiple receivers as well.

The received noise will be subject to the processing of the receiver. The output of a receiver z_i due to this noise is

$$z_i[n] = z_i(t)_{nT_s} = \left\{\int_{\bar{\mathbf{e}}\in\Omega_{\bar{\mathbf{e}}}} B_i(\bar{\mathbf{e}})\tilde{n}(t - \tau_i(\bar{\mathbf{e}}), \bar{\mathbf{e}})e^{-j\omega_o\tau_i(\bar{\mathbf{e}})}\,d\bar{\mathbf{e}} \ast \ast h(t)\right\}_{|nT_s} \qquad (15)$$

and has an expected value

11

$$E\{z_i[n]\} = 0 \qquad .\tag{16}$$

The cross covariance between the received noise for two sensors can be written

$$
\begin{aligned}
r_{ij} &= E\{z_i(t)z_j^*(t)\}\\[1ex]
&= E\left\{\left(\int_{\bar{\mathbf{e}}_1 \in \Omega_{\bar{\mathbf{e}}}} B_i(\bar{\mathbf{e}}_1)\tilde{n}(t - \tau_i(\bar{\mathbf{e}}_1),\bar{\mathbf{e}}_1)e^{-j\omega_o\tau_i(\bar{\mathbf{e}}_1)}\,d\bar{\mathbf{e}}_1 * *h(t)\right)\right.\\[1ex]
&\qquad \left.\left(\int_{\bar{\mathbf{e}}_2 \in \Omega_{\bar{\mathbf{e}}}} B_j(\bar{\mathbf{e}}_2)\tilde{n}(t - \tau_j(\bar{\mathbf{e}}_2),\bar{\mathbf{e}}_2)e^{-j\omega_o\tau_j(\bar{\mathbf{e}}_2)}\,d\bar{\mathbf{e}}_2 * *h(t)\right)^*\right\} \qquad , \tag{17}
\end{aligned}
$$

which can be shown to be

$$r_{ij} = \frac{4}{2\pi}\int_{\bar{\mathbf{e}} \in \Omega_{\bar{\mathbf{e}}}}\int_x^\infty |\mathcal{A}(\omega)|^2 B_i(\bar{\mathbf{e}})B_j^*(\bar{\mathbf{e}})\tilde{S}(\omega,\bar{\mathbf{e}})e^{j(\omega+\omega_o)(\tau_j(\bar{\mathbf{e}})-\tau_i(\bar{\mathbf{e}}))}\,d\bar{\mathbf{e}}\,d\omega \qquad . \tag{18}$$

where $\mathcal{A}(\omega)$ is the Fourier transform of the matched-filter impulse response.

Making the substitution of variables $\omega = (\omega - \omega_o)$ and defining

$$S(\omega,\bar{\mathbf{e}}) = 4\,\mathcal{A}(\omega - \omega_o)^2\tilde{S}(\omega - \omega_o,\bar{\mathbf{e}}) \qquad . \tag{19}$$

this can be written

$$r_{ij} = \frac{1}{2\pi}\int_{\bar{\mathbf{e}} \in \Omega_{\bar{\mathbf{e}}}}\int_x^\infty S(\omega,\bar{\mathbf{e}})B_i(\bar{\mathbf{e}})B_j^*(\bar{\mathbf{e}})e^{j\omega(\tau_j(\bar{\mathbf{e}})-\tau_i(\bar{\mathbf{e}}))}\,d\bar{\mathbf{e}}\,d\omega \qquad . \tag{20}$$

The argument of the exponential and the constant due to the sensor directional response can be expressed by components of a direction vector as

$$B_i(\bar{\mathbf{e}})B_j^*(\bar{\mathbf{e}})e^{j\omega(\tau_j(\bar{\mathbf{e}})-\tau_i(\bar{\mathbf{e}}))} = d_i(\omega,\bar{\mathbf{e}})d_j^*(\omega,\bar{\mathbf{e}}) \qquad . \tag{21}$$

and the covariance for the received data vector can be written as

$$\mathbf{R}[n] = E\{\mathbf{z}[n]\mathbf{z}^\dagger[n]\} = \frac{1}{2\pi}\int_{\bar{\mathbf{e}} \in \Omega_{\bar{\mathbf{e}}}}\int_x^\infty S(\omega,\bar{\mathbf{e}})\mathbf{d}(\omega,\bar{\mathbf{e}})\mathbf{d}^\dagger(\omega,\bar{\mathbf{e}})\,d\omega\,d\bar{\mathbf{e}} \qquad . \tag{22}$$

If the assumption that the system is narrowband is again made, the covariance matrix can be written as

12

$$\mathbf{R}(n) = \mathrm{E}\{\mathbf{z}(n)\mathbf{z}^\dagger(n)\} = \frac{1}{2\pi} \int_{\bar{\mathbf{e}} \in \Omega_{\bar{\mathbf{e}}}} S(\omega_o, \bar{\mathbf{e}}) \mathbf{d}(\omega_o, \bar{\mathbf{e}}) \mathbf{d}^\dagger(\omega_o, \bar{\mathbf{e}}) \, d\bar{\mathbf{e}} \qquad . \tag{23}$$

It will be assumed that the contribution of the spatially distributed noise sources to the received data vector can be modeled as a Gaussian distributed random variable. This model is based on the premise that the spatially distributed noise sources are Gaussian distributed or that there is a sufficient number of independent and identically distributed noise sources that the central limit theorem can be applied to model the resulting contribution as a Gaussian source.

The receiver-generated thermal noise can also be assumed to be subject to the Gaussian distribution with mean $\mathbf{0}$. The noise that is generated in each receiver is assumed to be independent with a contribution to the covariance of the sample vectors that will be a diagonal matrix

$$\mathbf{R}(n) = \mathrm{diag}(\sigma_1^2 \ldots \sigma_N^2) \qquad . \tag{24}$$

The output noise which is attributed to receiver noise sources is assumed to be independent of the spatially distributed noise sources. In high-performance receivers the noise statistics are very tightly controlled so that these receiver noise variances may be known. Additionally, the sources of noise in each of the receivers may be at the same level with the result that the individual receiver outputs are identically distributed. These additional pieces of information can be utilized in the beamforming and detection algorithms.

It is assumed that the in-phase and quadrature components of each of the receiver outputs are independent and identically distributed. The joint density function for the real and imaginary parts of the received vector will be written using the complex Gaussian density function.

$$f(\mathbf{z}; \mathbf{R}, b\mathbf{d}) = \frac{1}{\pi^N |\mathbf{R}|} e^{(\mathbf{z}-b\mathbf{d})^\dagger \mathbf{R}^{-1} (\mathbf{z}-b\mathbf{d})} \qquad . \tag{25}$$

The formulation of the complex density function in this manner comes from Goodman [5].

In summary, the time samples at the output of the receiver may consist of the sum of three terms:

$$\mathbf{z}(n) = b(n)\mathbf{d} + \mathbf{n}_s(n) + \mathbf{n}_r(n) \tag{26}$$

with mean

$$\mathrm{E}\{\mathbf{z}(n)\} = \begin{cases} b(n)\mathbf{d} & nT_s = \tau_o \\ \mathbf{0} & nT_s \neq \tau_o \end{cases} \qquad , \tag{27}$$

13

and with covariance given by the following expression

$$\mathbf{R}[n] = \frac{1}{2\pi} \int_{\bar{\mathbf{e}} \in \Omega_{\bar{\mathbf{e}}}} S(\omega_o, \bar{\mathbf{e}}) \mathbf{d}(\omega_o, \bar{\mathbf{e}}) \mathbf{d}^\dagger(\omega_o, \bar{\mathbf{e}}) \, d\bar{\mathbf{e}} + \text{diag}(\sigma_1^2 \ldots \sigma_N^2) \qquad . \qquad (28)$$

With a suitable choice of the modulation $A(t)$, the matched filtering in the receiver, and a suitable choice of the sampling interval T_s, the received data vectors can be assumed independent.

It is often valid to assume that targets are sparse: therefore, the number of samples which do not contain a target will be much greater than the number of samples which do. As a simplification of the data model. it will be assumed that the target signal energy will appear in the mean of only one sample for each transmitted pulse and this pulse is known. All other samples are assumed to be zero mean. This model, with the target energy assumed to be indicated by the mean of the received data. is the standard non-fluctuating model for radar targets [6,4].

A simplifying change of notation will be made here, and the temporal samples $\mathbf{z}[n]$ will be denoted

$$\mathbf{z}_n \equiv \mathbf{z}[n] \qquad . \qquad (29)$$

The sampled data vectors will be arranged into a data matrix $\mathbf{Z}$. Column n of this matrix is $\mathbf{z}_n$. The data can be ordered so that returns that (may) contain a target return. $nT_s = \tau_o$, are the first G vector samples corresponding to G transmitted pulses. These data vectors will be called the *primary* data vectors. When there is only a single primary vector, $G = 1$, and there is little chance of misunderstanding, the primary vector will be separated from the data matrix and will be denoted $\mathbf{z}$. The next K of the samples are assumed to have zero mean. corresponding to data vectors that do not contain a target return. These data vectors will be called the *secondary* data vectors.

The mean of $\mathbf{Z}$ can be expressed as

$$\mathbf{B} \equiv \mathrm{E}(\mathbf{Z}) \qquad . \qquad (30)$$

The first G columns of $\mathbf{B}$ will be complex multiples of the known direction vector, i.e., $b_i\mathbf{d}$, and the remaining K columns will be zero.

The columns of $\mathbf{Z}$ are assumed to be mutually independent with covariance

$$\text{Cov}(\mathbf{z}_i) = \mathbf{R} \qquad . \qquad (31)$$

14

2.5 Coordinates

In the above derivation, an explicit definition of the coordinate system has been unnecessary. For many interference environments, the model lends itself well to using spherical coordinates for the integral equation, as shown in Figure 5. The area differential $d\bar{e}$ is $\sin\theta d\theta d\phi$ and Equation (28)
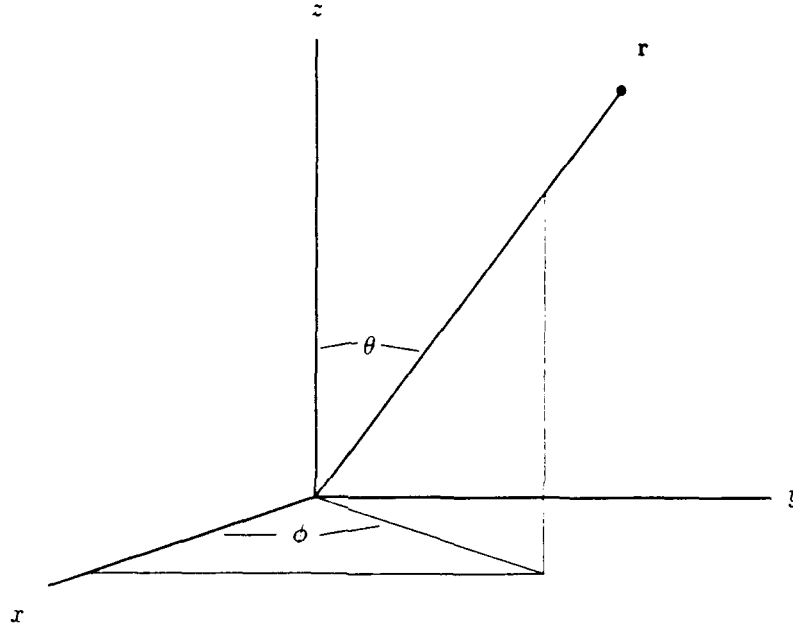
Figure 5. Coordinate system.

can be expressed

$$\mathbf{R}_{[n]} = \frac{1}{2\pi} \int_{\theta \in \Theta, \phi \in \Phi} S(\omega_o, \bar{\mathbf{e}}) \mathbf{d}(\omega_o, \theta, \phi) \mathbf{d}^{\dagger}(\omega_o, \theta, \phi) \sin\theta \, d\theta d\phi + \text{diag}(\sigma_1^2 \ldots \sigma_N^2),$$ (32)

where the regions of integration have been defined as Θ and Φ.

15

2.6 Communications Model

For a communications system, the receiver is passive, and it is assumed that the transmitter transmits a signal of the form

$$s(t) = \sqrt{E_t}\, \Re\{b(t)A(t)e^{j\omega_o t}\} \qquad . \tag{33}$$

The transmitted information is contained in the amplitude and phase of the modulation term b. This signal may change from sample to sample according to the transmitted modulation. This yields a model for the received data that is nearly the same as that for a radar system. In the radar system, targets are assumed to be sparse, and there will be only a few columns of $\mathbf{Z}$ that are non-zero mean. In a communications system many or all of the columns of $\mathbf{Z}$ may be non-zero mean.

2.7 Examples

2.7.1 Spatially Uniform Interference

A two-sensor example will be used in order to illustrate the covariance that results from the integral expression found in Equation (32) for the covariance matrix of the spatially distributed noise. The sensors are assumed to have omnidirectional gain which can be arbitrarily set to unity. The interference is assumed to be distributed uniformly over all directions of arrival. A spherical coordinate system will be utilized, and the coordinate system will be rotated such that the sensors will lie on the z axis, and the term $\bar{\mathbf{e}} \cdot (\mathbf{s}_i - \mathbf{s}_j)$ becomes $|\mathbf{s}_i - \mathbf{s}_j| \cos\theta$. This is shown in Figure 6. The spatial-temporal spectral-density function will be denoted by $S(\omega_o)$ since it is independent of direction. The elements of Equation (32) can then be written

$$r_{ij} = \frac{S(\omega_o)}{2\pi} \int_{\theta \in \Theta, \phi \in \Phi} e^{j\frac{\omega_o}{c}|\mathbf{s}_i - \mathbf{s}_j|\cos\theta} \sin\theta\, d\theta\, d\phi \qquad . \tag{34}$$

Performing this integration results in the expression

$$r_{ij} = 2S(\omega_o) \frac{\sin\frac{\omega_o}{c}|\mathbf{s}_i - \mathbf{s}_j|}{\frac{\omega_o}{c}|\mathbf{s}_i - \mathbf{s}_j|} \qquad . \tag{35}$$

This then describes the correlations between sensors as a function only of the distance between the sensors, the propagation time of the medium, and ω_o. When the distance between sensors is a multiple of one-half wavelength, $|\mathbf{s}_i - \mathbf{s}_j| = n\frac{\pi c}{\omega_o}$, the interference will be uncorrelated between two sensors, and the resulting covariance matrix will be a multiple of the identity matrix.
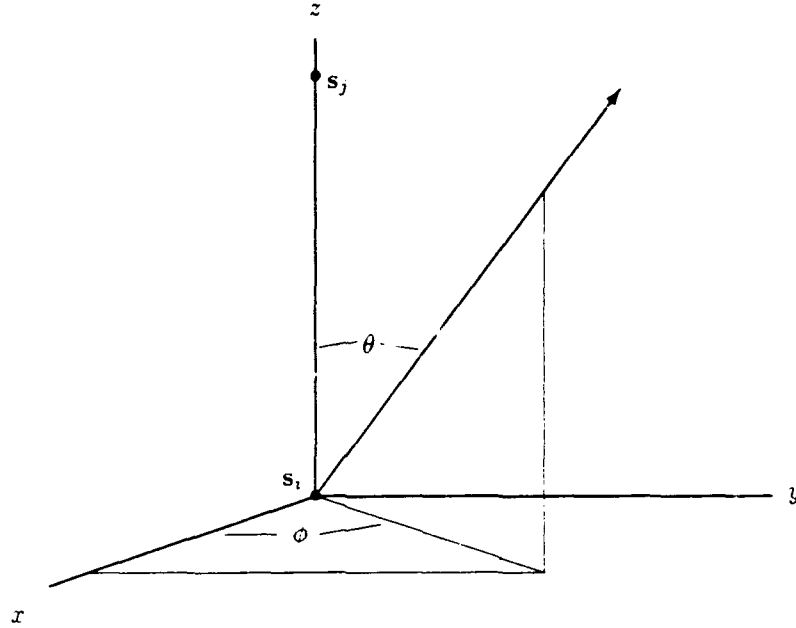
Figure 6. Coordinate system for two sensor "linear" array.

2.7.2 Toeplitz Structure

It is commonly known that uniform linear sampling of a spatially homogeneous noise field results in a Toeplitz correlation matrix. This is given in Equation (20), which is repeated here for convenience.

$$r_{ij} = \frac{1}{2\pi} \int_{\bar{e} \in \Omega_{\bar{e}}} \int_{\infty}^{\infty} S(\omega, \bar{e}) B_i(\bar{e}) B_j^*(\bar{e}) e^{j\omega(\tau_j(\bar{e}) - \tau_i(\bar{e}))} \, d\bar{e}d\omega \qquad . \tag{36}$$

There are two requirements that must be satisfied in order to have uniform sampling. The first is that the antenna spatial responses $B_i(\bar{e})$ must be the same for all i. Additionally, with suitable scaling, the term $(\tau_j(\bar{e}) - \tau_i(\bar{e})) = \bar{e} \cdot (s_i - s_j)$ must be expressible as a function of $i - j$. The resulting covariance matrix has elements that are a function only of $i - j$ and then has the Toeplitz structure.

17

The results of this report can be applied to detection and beamforming problems where the true covariance is assumed to be Toeplitz; however, the intent of this report is to provide methods that can be utilized for arbitrary sensor geometries and spatial responses.

2.8 Problem Statements

2.8.1 Detection

The purpose of detection is to determine if a target is present or not. Target presence is indicated by a non-zero mean in a single data sample. Additional target-free data are available from a process having the same noise covariance matrix. The direction toward a possible target is known and has a corresponding direction vector $\mathbf{d}$.

This is a binary hypothesis testing problem, and the detector will be designed with the intent of satisfying the Neyman-Pearson criterion. The two hypotheses and the data that are available under each hypothesis are shown in Table 1. This data model will be specialized in some of the

TABLE 1

Detection Data Model

	Hypothesis	Data	Density
H_0	No target present	$\mathbf{z}_1 \ldots \mathbf{z}_{K+G}$	N(0, $\mathbf{R}$)
H_1	Target present	$\mathbf{z}_1 \ldots \mathbf{z}_G$	N($b\mathbf{d}$, $\mathbf{R}$)
		$\mathbf{z}_{G+1} \ldots \mathbf{z}_{K+G}$	N(0, $\mathbf{R}$)
where $\mathbf{R} = \int_{\bar{\mathbf{e}} \in \Omega_{\bar{\mathbf{e}}}} S(\bar{\mathbf{e}})\mathbf{d}(\bar{\mathbf{e}})\mathbf{d}^{\dagger}(\bar{\mathbf{e}}) \, d\bar{\mathbf{e}} + diag[\sigma_1^2 \cdot \sigma_N^2]$			

sections of this report. Restrictions will be placed on the number of terms K and G so that a detector based on maximum-likelihood techniques will exist.

It will be assumed that the mean b and covariance $\mathbf{R}$ are unknown. The true covariance is assumed to be non-singular. The detection results for the methods that are presented in this report will be compared to the performance of detectors where $\mathbf{R}$ is known.

18

2.8.2 Beamforming

The beamforming problem is a parameter estimation problem where the parameter of interest is the signal mean(s) $b[n]$. The direction of arrival of the transmitted waveform and the corresponding direction vector $\mathbf{d}$ are known. The data that are available are shown in Table 2. The covariance

TABLE 2

Beamformer Data Model

Data	Density
$\mathbf{z}_1 \ldots \mathbf{z}_G$	$N(\; b[n]\mathbf{d},\; \mathbf{R})$
$\mathbf{z}_{G-1} \ldots \mathbf{z}_{K-G}$	$N(\; 0,\; \mathbf{R})$
where $\mathbf{R} = \int_{\bar{\mathbf{e}} \in \Omega_{\bar{\mathbf{e}}}} S(\bar{\mathbf{e}})\mathbf{d}(\bar{\mathbf{e}})\mathbf{d}^\dagger(\bar{\mathbf{e}})\, d\bar{\mathbf{e}} + diag[\sigma_1^2 \cdot \sigma_N^2]$	

matrix is assumed to be unknown and non-singular. The estimates produced by the methods which are presented in this report will be compared to the estimates produced when $\mathbf{R}$ is known.

3. CURRENT METHODS

3.1 Beamforming

Often the complex envelope of a propagating wave contains desirable information which is to be extracted by signal processing. When the environment consists of interference sources whose temporal frequency content overlaps that of the desired signal, then temporal filtering alone cannot always be used to separate the interference from the desired signal. If the interference and the desired signal arrive at the sensors from different spatial angles, the angular separation can also be exploited to extract the desired signal from the interference. This spatial filtering process is called beamforming.

Spatial filters process data over a non-zero spatial aperture. This processing can be provided by an antenna with a continuous spatial aperture, such as a reflector or lens antenna, or it can be provided by combining discrete spatial samples and beamforming.

For many continuous aperture antennas, the desired spatial response is a pencil beam where the response is limited to a small angular region about the direction of the desired signal. This response region is known as the *main lobe*, and other typically undesirable responses are known as the *sidelobes*. The term "beamforming" when used with spatial arrays originates with this connection to continuous aperture antennas [7]. This connotation of forming beams may be misleading since for "optimum" beamformers the desired response may not be a pencil beam.

There are many methods that have been utilized for combining the outputs of multiple sensors to realize beamformers. One method is to delay and sum the outputs of the sensors so that wavefields propagating from a desired direction will constructively add. This delay and sum beamformer is applicable to either narrowband or wideband signals. For narrowband wavefields and receivers, it may be implemented by phase shifting rather than delaying the received signals. These methods typically utilize analog processing of the sensor data; however, they are applicable to the processing of the receiver outputs as well. This report will focus on methods of combining the samples of narrowband receiver outputs.

Beamformers are typically a linear function of the sensor array outputs. This equation can be written as

$$\hat{b} = \mathbf{w}^\dagger \mathbf{z} \quad , \tag{37}$$

where the weight vector $\mathbf{w}$ indicates the linear combination of the array data $\mathbf{z}$, and $\hat{b}$ is the estimate of the complex envelope. Often, conventional beamformers for antenna arrays attempt to form a pencil beam similar to continuous aperture antennas and typically utilize a fixed weight vector. This is used to provide the resolution advantages of a large array without incurring the mechanical disadvantages. The performance of these approaches is limited by the spatial response of the resulting beamformers. Interference that appears in the sidelobes of the beam pattern can

mask the desired signal. Several approaches, both ad hoc and optimal, have been proposed that partially alleviate the sidelobe problem at the expense of main lobe width. Since these methods do not take the interference environment into account, strong interference can still obscure the desired signal, resulting in an unacceptable error in the estimate. Performance criteria can be used to derive estimators which account for the interference.

One performance criterion which can be used is to form an estimate of b that minimizes the mean-squared error:

$$\mathbf{w}_o = \underset{\mathbf{w}}{\text{argmin }} \mathrm{E}\{|b - \mathbf{w}^\dagger \mathbf{z}|^2\}, \quad \hat{b} = \mathbf{w}_o^\dagger \mathbf{z} \qquad . \tag{38}$$

This method was investigated by Wiener for scalar continuous time systems [8]. Levinson [9] reformulated the minimum mean-square error problem for discrete systems and developed the normal equations for finding a transversal filter or colored-noise matched filter that minimizes the mean-square error between the estimate and the true signal. This estimator requires that the noise covariance be known. If the signal mean is of the form $b\mathbf{d}$ and the noise covariance is $\mathbf{R}$, the resulting optimal weight vector is

$$\mathbf{w}_o = k\mathbf{R}^{-1}\mathbf{d} \tag{39}$$

with $k = |b|^2/(1 + |b|^2 \mathbf{d}^\dagger R^{-1} \mathbf{d})$ [7]. The resulting *estimate* of b is biased and minimizes the mean-square error for a given value of b. The results of Levinson apply directly to spatial filtering. The response of the spatial filter that minimizes the mean-square error will tend to put nulls in the magnitude of the response in directions corresponding to strong interference sources [7].

Another performance criterion is the maximization of the signal-to-noise ratio at the output of a linear filter [10]. Defining the expected values E_N and E_{S+N} as the expected values when only noise is present or when signal and noise are present, the signal-to-noise ratio at the output of a linear filter is [2]

$$SNR = \frac{|E_{S+N}\{\mathbf{w}^\dagger \mathbf{z}\} - E_N\{\mathbf{w}^\dagger \mathbf{z}\}|^2}{E_N\{|\mathbf{w}^\dagger \mathbf{z}|^2\}} \qquad . \tag{40}$$

For our problem where the output is zero mean when the input is assumed to be noise-only, the signal-to-noise ratio is [10]

$$SNR = |b|^2 \frac{|\mathbf{w}^\dagger \mathbf{d}|^2}{\mathbf{w}^\dagger \mathbf{R} \mathbf{w}} \qquad . \tag{41}$$

22

The linear filter which maximizes the signal-to-noise ratio is given by [10]

$$\mathbf{w}_o = k\mathbf{R}^{-1}\mathbf{d} \qquad , \qquad (42)$$

where k is an arbitrary constant. The Wiener filter, which minimizes the mean-square error, will then also maximize the output signal-to-noise ratio. For the optimal linear filter, the resulting signal-to-noise ratio is [10]

$$SNR_o = |b|^2 \mathbf{d}^\dagger \mathbf{R}^{-1}\mathbf{d} \qquad . \qquad (43)$$

This is the signal-to-noise ratio that would be achievable if the noise statistics were known. This signal-to-noise ratio appears repeatedly in detector performance analysis, so it is convenient to define the following:

$$a \equiv SNR_o \qquad . \qquad (44)$$

Capon [11] proposed a similar problem where it is desired to minimize the power out of a linear filter while constraining the gain to unity in the desired reception direction. The resulting beamformer is known as the minimum-variance distortionless-response (MVDR) beamformer, and the weight vector is given by [11] [2]

$$\mathbf{w} = \frac{\mathbf{R}^{-1}\mathbf{d}}{\mathbf{d}^\dagger \mathbf{R}^{-1}\mathbf{d}} \qquad . \qquad (45)$$

The covariance matrix again must be known to evaluate this expression. The variance of the estimates produced by this estimator is [11]

$$E\{|\mathbf{w}^\dagger \mathbf{z} - E\{\mathbf{w}^\dagger \mathbf{z}\}|^2\} = \frac{1}{\mathbf{d}^\dagger \mathbf{R}^{-1}\mathbf{d}} \qquad . \qquad (46)$$

This function, expressed as a function of θ (where $\mathbf{d} = \mathbf{d}(\theta)$) using the estimate of the covariance to be introduced in Equation (48), is known as Capon's spectral estimator [12].

---

[2]The weight vector that is derived by this method uses the correlation matrix $\mathbf{R} + |b|^2\mathbf{d}\mathbf{d}^\dagger$ in this formula rather than the covariance matrix $\mathbf{R}$. By application of the matrix inversion lemma, $(\mathbf{R} + |b|^2\mathbf{d}\mathbf{d}^\dagger)^{-1} = \mathbf{R}^{-1} - |b|^2\mathbf{R}^{-1}\mathbf{d}\mathbf{d}^\dagger\mathbf{R}^{-1}/(1 + |b|^2\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{d})$, these two forms are equivalent.

The methods discussed to this point have not made use of the density function for the received data. Using the Gaussian model derived in Chapter 2 and assuming that the noise covariance is known, the maximum-likelihood estimate of the mean term is given by the formula [12]

$$\hat{b} = \frac{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{z}}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d}} \ . \tag{47}$$

This estimate of b also results from using the MVDR beamformer on the data sample $\mathbf{z}$. For this reason Equation (46) is sometimes called the ML spectrum estimator [12]. The Cramer-Rao bound on the variance of an unbiased estimate of b is given by Equation (46), and thus, the MVDR beamformer is an efficient estimator of b.

Knowledge of the covariance matrix has been used in the design of the beamformers discussed thus far. The noise covariance matrix is seldom known, leading to adaptive approaches where the weight vector or noise covariance is estimated as well as the mean. There have been many methods proposed to solve iteratively for the optimal minimum mean-square error weight vector as the data are received [7]. The performance of these methods in terms of the resulting signal-to-noise ratio is difficult to analyze, especially during adjustment. For stochastic gradient approaches, the rate of convergence is dependent upon a step size parameter and the distribution of eigenvalues of the true covariance matrix. Additionally, there may be an increase in the output signal-to-noise ratio even as the number of samples applied for adaptivity approaches infinity. These methods approximate, in an iterative fashion, the sample matrix inversion method that will be discussed next. The iterative approaches will not be discussed further.

Reed, Mallett, and Brennan (RMB)[10], using the results of Goodman [5], proposed using the maximum-likelihood estimate of the covariance matrix to form a matched-filter beamformer that adapts to the interference environment. This was one of the first papers to propose using statistical methods based on all of the data that are present to estimate the weight vector. This method is based on the model where K mutually independent and identically distributed data samples, assumed to be zero mean, are available and which can be utilized to form an estimate of the covariance matrix. The covariance estimate used is the average of the sum of outer products of noise-only samples:

$$\hat{\mathbf{R}} = \left(\frac{1}{K}\right) \mathbf{Z} \mathbf{Z}^\dagger \ . \tag{48}$$

The data samples are the columns of $\mathbf{Z}$ in this notation. This estimate is the unconstrained maximum-likelihood estimate in that it does not incorporate any of the covariance modeling of Chapter 2. This estimate is substituted into the known covariance matched filter to form the weight vector

$$\mathbf{w} = k\hat{\mathbf{R}}^{-1}\mathbf{d} \ , \tag{49}$$

24

where k is an arbitrary constant.

The ratio of the signal-to-noise ratio at the output of a linear beamformer to the signal-to-noise ratio of the optimal linear beamformer is given by the equation

$$\rho = \frac{|\mathbf{w}^\dagger \mathbf{d}|^2}{\mathbf{d}^\dagger \mathbf{R}^{-1}\mathbf{d}\,\mathbf{w}^\dagger \mathbf{R}\mathbf{w}}$$

$$= \frac{|\mathbf{d}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{d}|^2}{\mathbf{d}^\dagger \mathbf{R}^{-1}\mathbf{d}\,\mathbf{d}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{d}} \quad . \tag{50}$$

When the unconstrained maximum-likelihood estimate of the covariance is used to form the weight vector, this signal-to-noise ratio loss factor is a random variable whose statistics can be described in terms of the dimensional parameters N (number of sensors) and K (number of samples). The loss factor in this case is beta distributed over the range $[0,1]$ with a density given by [10]

$$f_3(\rho) = \frac{K!}{(N-2)!(K+1-N)!}(1-\rho)^{N-2}\rho^{K+1-N} \quad , \tag{51}$$

with an expected value of $\frac{K+2-N}{K+1}$.

The original paper by Reed et al. [10] proposed using the beamformer to perform detection. It has been common to use their suggestion of replacing the unknown true covariance with the maximum-likelihood estimate of the covariance in adaptive beamformers to estimate the complex modulation. Adaptive beamforming is performed on non-zero mean data using a covariance matrix estimate based on zero mean data. When this method is used with Capon's MVDR beamformer, it can be shown that the output for a single sample is the maximum-likelihood estimate of b found via the joint maximization of the likelihood function over the mean and the covariance matrix.

One problem with this use of the maximum-likelihood estimate of the noise covariance to form the weight vector is that there can be a large loss in signal-to-noise ratio for a small number of vector samples forming the estimate. If the number of vector samples is too small ($K < N$), then the sample covariance matrix will be rank deficient and the inverse will not exist.

3.2 Detection

Adaptive beamforming and detection have historically been combined due to their close relationship. A common approach to detection has been to form a beam in a particular direction to get an estimate of the mean in that direction and then to compare the magnitude of this estimate to a fixed threshold to determine if the signal is present. The probability of false alarm for this test will vary with the noise covariance, so detectors based on this method do not have the constant false alarm rate (CFAR) property. Various adaptive approaches to setting the threshold have been proposed. Many of these utilize some type of "cell-averaging" or cell comparison where the estimated signal power ($|\hat{b}|^2$) for the vector under test (the test cell) is compared to a combination

25

of the estimated signal powers for other vector samples and for different directions (the reference cells) [3]. The resulting detectors can provide CFAR behavior. These approaches are justified by the observation that the level of noise from the reference cells would be approximately the same as that of the cell under test. If the beam that is formed to estimate the mean does not adapt to the interference environment, then strong interference in the sidelobes of the array response can result in a poor signal-to-noise ratio and poor detector performance.

A Neyman-Pearson optimal detector is based on the likelihood ratio test (LRT) [2,13]. The likelihood ratio test is

$$\Lambda(\mathbf{z}) = \frac{l_{H_1}(\mathbf{z})}{l_{H_0}(\mathbf{z})} \overset{H_1}{\underset{H_0}{\gtrless}} \gamma \qquad . \tag{52}$$

Using the discussed techniques [2], the test for this problem can be written as

$$\frac{\mathrm{Re}(b^* \mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{z})}{\sqrt{(|b|^2 \mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d})}} \overset{H_1}{\underset{H_0}{\gtrless}} \gamma \qquad . \tag{53}$$

The normalization constants in the denominator have been introduced to simplify the expressions for the probability of false alarm and the probability of detection. By using various methods [10,2], the probability of false alarm for this test is

$$PFA = \frac{1}{2} \mathrm{erfc}\,(\gamma) \qquad , \tag{54}$$

and the probability of detection is

$$PD = \frac{1}{2} \mathrm{erfc}\,(\gamma - a) \qquad . \tag{55}$$

a is defined in Equation (44), and the complementary error function is given by the expression

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt \qquad . \tag{56}$$

This detector is not practical because the true covariance $\mathbf{R}$ and the mean b are rarely known.

A similar test, assuming that the mean is unknown and the noise covariance is known, is derived using the generalized likelihood ratio test (GLRT). The GLRT requires a maximization of

26

the likelihoods over the unknown parameters Θ of the two hypotheses before forming the likelihood ratio [2]:

$$\Lambda(\mathbf{z}) = \frac{\max_{\Theta \ H_1} l_{H_1}(\mathbf{z};\Theta)}{\max_{\Theta \ H_0} l_{H_0}(\mathbf{z};\Theta)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \qquad . \tag{57}$$

The test derived by this method is

$$\frac{|\mathbf{d}^\dagger \mathbf{R}^{-1}\mathbf{z}|^2}{\mathbf{d}^\dagger \mathbf{R}^{-1}\mathbf{d}} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \qquad . \tag{58}$$

This test is not uniformly most powerful because the LRT in Equation (53) can be used assuming a particular value for the mean, and the probability of detection will be higher when the true mean is near the design value. The generalized likelihood ratio test of Equation (58) can be written as the ratio of the signal power estimated by the maximum-likelihood method to the MVDR spectral estimate

$$\frac{|\hat{b}|^2}{1/\mathbf{d}^\dagger \mathbf{R}^{-1}\mathbf{d}} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{59}$$

and is the test that results when the maximum-likelihood estimate of b is substituted without the normalization in the denominator into Equation (53). This is an example of the relationship between beamforming (estimation of b) and detection.

Other methods [10.2] can be used to determine the detection performance. The probability of false alarm for this test is

$$PFA = e^{-\gamma} \qquad . \tag{60}$$

and the probability of detection is

$$PD = Q(\sqrt{2a}, \sqrt{2\gamma}) \qquad . \tag{61}$$

where Q is the Marcum's Q-function

$$Q(a,b) = \int_b^{\infty} u e^{-\frac{u^2 + a^2}{2}} I_o(au) du \qquad . \tag{62}$$

The performance of detection algorithms that are proposed in this report will be compared to the performance of these known parameter detection algorithms.

3.2.1 Unknown Noise Covariance GLRT

Kelly [14,15] has derived a generalized likelihood ratio detector based on the model that K mutually independent zero mean secondary data vectors (the columns of Z) and a single, possibly non-zero mean, primary data vector are available. These data are assumed to share the same covariance matrix and are subject to the Gaussian density function. The unconstrained maximum-likelihood estimate of the covariance is used in an adaptive detection rule that is given for data sample z and threshold γ by [15]

$$\frac{|\mathbf{d}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{z}|^2}{\mathbf{d}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{d}\left(1 + \frac{1}{K}\mathbf{z}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{z}\right)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{63}$$

with

$$\hat{\mathbf{R}} = \frac{1}{K}\mathbf{Z}\mathbf{Z}^\dagger \qquad . \tag{64}$$

This detector requires an increase in signal-to-noise ratio to achieve a given probability of detection at a fixed probability of false alarm when compared to the known covariance detector. The required signal-to-noise ratio increase is greater than the loss due to adaptive beamforming; the additional loss is attributed to the determination of the threshold. The loss is independent of the true covariance, depending only upon the dimensional parameters and the probability of false alarm. The generalized likelihood ratio test results in a test in which the probability of false alarm is independent of the level and structure of the noise covariance, rather than just the "local" noise level as in the previous cell-averaging techniques.

We [16,17] have derived an adaptive detector which is simplified compared to the generalized likelihood ratio detector. This decision rule is defined by substituting the maximum-likelihood estimate of the covariance based on the zero mean data samples in place of the known covariance in the unknown mean detector of Equation (58). The derivation and analysis of this test are discussed further in Chapter 4. The detection rule.

$$\frac{|\mathbf{d}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{z}|^2}{\mathbf{d}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{d}} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \qquad . \tag{65}$$

28

compares the output power of a normalized matched filter to a threshold. This procedure is called the adaptive matched-filter or AMF detector. It will be shown that the AMF detector has the constant false alarm rate (CFAR) property similar to the generalized likelihood procedure. The probability of detection of this detector can show either an increase or decrease in detector performance compared to Kelly's generalized likelihood ratio detector: the difference in performance depends upon the value of the signal-to-noise ratio and the dimensional parameters.

Both of the detectors that use the unconstrained maximum-likelihood estimate of the covariance matrix suffer from problems that are similar to those of the unconstrained maximum-likelihood-based beamformers. A large increase in signal-to-noise ratio may be required in order to reach the same probability of detection as the test where the covariance is known. Furthermore, the tests will not exist if $K < N$ since the sample covariance matrix is singular.

3.3 Covariance Estimators

Although the covariance matrix is a nuisance parameter, it plays an essential role in the detection and beamforming algorithms. We will utilize the model for the covariance derived in Chapter 2 and estimate covariance matrices that are restricted to this model in the hopes that by estimating a matrix that is "closer" to the true covariance matrix, higher performance will result compared to estimates that do not make full use of the model. Estimators based on some of the aspects of the model in Chapter 2 have been used previously for some special arrays and interference environments: these will be discussed in this section.

3.3.1 Unconstrained Maximum Likelihood

The unconstrained maximum-likelihood estimate of the covariance causes several problems when used in adaptive beamformers. The main problem is that although the expected loss in signal-to-noise ratio may be small for a given number of samples used to estimate the covariance (on the order of $2N$ samples, the expected loss is 3 dB [10]) the variance of the loss is high. Any one realization of the adaptive beamformer may have an unacceptable loss in signal-to-noise ratio [18]. The variance of the estimate also generates an instability in the spatial response of the adaptive beamformer. This instability has been attributed to the sensitivity of the inverse of the estimated covariance matrix to the smallest eigenvalues [19,20] and is a function of the condition number of the covariance matrix.

The AMF and the generalized likelihood ratio detectors also suffer from poor sidelobe performance [16,21]. If the primary data vector has a source of interference that is not reflected in the other data samples either due to non-stationarity of the interference or because the sample contains an outlier, then detection performance will be poor. The AMF test has poor rejection of signals or interference appearing in the sidelobes of the range bin under test but which are not in the secondary data [16]. The effect of this is that a strong outlier will cause the detector to indicate that there is a signal in all directions, masking the detection of the signal which might appear in the same vector sample. For Kelly's GLRT detector, a single interferer in the primary vector will

cause the test statistic to be reduced for directions other than in the direction of the strong signal, and detection of the desired signal may not be accomplished [21]. The impact of the outliers is reflected in the probabilities of detection and false alarm; non-stationarity of the noise environment is not reflected.

3.3.2 Subspace Approaches

An attempt to improve the sidelobe performance of the adaptive beamformer over that of the unconstrained maximum-likelihood approach is based on modeling the interference as a number ($J < N$) of narrowband spatial point sources. The covariance is then given by the rank J term due to the emitters plus a diagonal term attributed to receiver noise. One method that uses this technique is based on the eigenvector decomposition of the sample covariance matrix [13,22]. This method assumes that the J largest eigenvalues correspond to the emitters and that the smallest $N - J$ eigenvalues correspond to white noise. The smallest eigenvalues are averaged (or set to a known value) to form the covariance estimate, which is then used in an adaptive beamformer. This is the maximum-likelihood estimate of $\mathbf{R}$ under this low-rank model.

Another method assumes that the emitter locations are known and then estimates a covariance matrix which is constrained to be of this form [23,24].

These methods are sensitive to errors in the subspace chosen for the discrete emitters and in the eigenvalue spread of the true covariance.

3.3.3 Covariance Modeling

Covariance models can be derived based on the array geometry and the spatial structure of the noise. A covariance estimate based on this model can then be utilized in an adaptive beamformer or detector. A common assumption is that the interference is located in the far field and is spatially independent. This assumption leads to a spatially stationary noise environment. A uniform linear array sampling this field results in a Toeplitz covariance matrix. Estimators for covariances that are constrained to have the Toeplitz structure are popular [25,26,27,28]. The two predominant methods of forming an estimate that has the Toeplitz structure are to use maximum-likelihood methodology or to use diagonal averaging.

Diagonal averaging minimizes the Frobenious norm of the difference between the sample covariance and the estimate. An unbiased estimate may result, but it will not always be positive-definite. Modifications have been introduced to ensure a positive-definite estimate.

The Toeplitz constrained maximum-likelihood estimate of the covariance has been utilized to form an adaptive detection algorithm [29], showing a dramatic increase in the probability of detection compared to the unconstrained case. This estimator is useful for uniform linear arrays; however, other array geometries do not lead to the Toeplitz structure.

The focus of this report is to improve detector and beamformer performance for arbitrary but known arrays and for an arbitrary number of interference sources. The results of this report are

not restricted to uniform linear arrays with the corresponding Toeplitz covariance matrix structure. Arrays with arbitrary spatial geometries can also lead to structured covariance matrices, although the structure is not as obvious as the Toeplitz structure.

This structure will be discussed further in Chapter 5. In the next chapter the CFAR adaptive matched-filter detector will be discussed.

4. AN ADAPTIVE MATCHED-FILTER DETECTOR

In this chapter, an adaptive detector is derived and the performance analyzed. This detector utilizes the Gaussian density function and the structured mean but does not utilize any information about the covariance structure. In order to form an unstructured estimate of the covariance, it is assumed that there are K independent signal-free data vectors that are used to form an unconstrained estimate of the covariance matrix. This estimate is then substituted into the known covariance generalized likelihood ratio test. The resulting test statistic has the form of a normalized colored-noise matched filter and is a CFAR detector.

The purpose in investigating this detector is that the resulting test statistic is simpler than that derived through the full unconstrained generalized likelihood ratio procedure. For real-time applications, the difference in complexity may be significant. Adaptive detectors using constrained covariance estimates can be derived by the same method used to derive the detector presented in this chapter. This detector and the unconstrained generalized likelihood ratio detector provide a performance base to which the performance of constrained adaptive detectors can be compared.

Equations describing the performance of this detector are derived for signals on boresight (i.e., in the $\mathbf{d}$ direction) as well as for signals which are not matched to boresight. The performance of this detector will be compared in detail to other detection methods in Chapter 9.

4.1 Derivation of the Test Statistic

The signal model is a slight variation of that given in Chapter 2. For notational simplicity the *primary* data vector $\mathbf{z}$ will be considered separately from the *secondary* data. $\mathbf{z}$ is assumed to be a complex N-length Gaussian random vector with mean $\mathbf{0}$ under hypothesis H_0, mean $b\mathbf{d}$ under hypothesis H_1, and positive-definite covariance matrix $\mathbf{R}$. $K > N$ additional independent data vectors are arranged as the columns of the data matrix $\mathbf{Z}$. These secondary data vectors are assumed to have mean $\mathbf{0}$ and covariance $\mathbf{R}$. The restriction on the number of secondary data vectors is made so that the maximum-likelihood estimate of the covariance matrix will exist almost surely.

The procedure used to derive the test statistic is to assume that the covariance is known and then to write the generalized likelihood ratio test maximizing over the unknown parameter b. The resulting test statistic is the output power of the standard colored-noise matched filter. The maximum-likelihood estimate of the covariance based on the secondary data alone is then substituted into this test.

The derivation is begun by writing the generalized likelihood ratio test

$$\Lambda = \frac{\max_b f_{\mathbf{z}|H_1}(\mathbf{z}; b|H_1)}{f_{\mathbf{z}|H_0}(\mathbf{z}|H_0)} \overset{H_1}{\underset{H_0}{\gtrless}} \gamma \quad . \tag{66}$$

33

Substituting the complex multivariate Gaussian density functions and canceling common terms yields

$$\Lambda = e^{-(\mathbf{z}-b\mathbf{d})^\dagger \mathbf{R}^{-1}(\mathbf{z}-b\mathbf{d})+\mathbf{z}^\dagger \mathbf{R}^{-1}\mathbf{z}} \quad . \tag{67}$$

Now the logarithm can be taken, and the result simplified to

$$\log(\Lambda) = 2\Re(b^* \mathbf{d}^\dagger \mathbf{R}^{-1}\mathbf{z}) - |b|^2 \mathbf{d}^\dagger \mathbf{R}^{-1}\mathbf{d} \quad . \tag{68}$$

Maximizing this with respect to the unknown complex amplitude b yields

$$\hat{b} = \frac{\mathbf{d}^\dagger \mathbf{R}^{-1}\mathbf{z}}{\mathbf{d}^\dagger \mathbf{R}^{-1}\mathbf{d}} \quad . \tag{69}$$

Substituting Equation (69) into Equation (68) and simplifying produces the test

$$\frac{|\mathbf{d}^\dagger \mathbf{R}^{-1}\mathbf{z}|^2}{\mathbf{d}^\dagger \mathbf{R}^{-1}\mathbf{d}} \underset{H_0}{\overset{H_1}{\gtrless}} \alpha \quad , \tag{70}$$

with $\alpha = \log(\gamma)$. This test statistic is proportional to the squared magnitude of the output of the colored-noise linear matched filter because the term in the denominator is a constant when the true covariance is known.

If the noise covariance matrix were known, then the detector described by Equation (70) would be used. In general, the covariance matrix is unknown and must be accounted for by using adaptive techniques. The generalized likelihood ratio test (GLRT) provides one such adaptive approach. We propose to account for not knowing the true covariance by the ad hoc procedure of substituting the maximum-likelihood estimate based on the secondary data.

$$\hat{\mathbf{R}} = \frac{1}{K}\mathbf{Z}\mathbf{Z}^\dagger \quad . \tag{71}$$

Reed, Mallett, and Brennan (RMB) used a similar approach in their maximum signal-to-noise formulation of the detection problem. The test form is then

$$\frac{|\mathbf{d}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{z}|^2}{\mathbf{d}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{d}} \underset{H_0}{\overset{H_1}{\gtrless}} \alpha \quad . \tag{72}$$

34

We call this test the adaptive matched-filter (AMF) test. This test statistic has the RMB test statistic as the numerator, with a normalization that is the same as that which would be provided by the unconstrained GLRT for a large number of secondary samples. This normalization will provide the desired CFAR behavior and is a natural normalization factor to use for this purpose.

The AMF test may also be derived by other methods. This test statistic also results from a type of cell-averaging CFAR where the cell average is made from the outputs of an RMB adaptive beamformer [16].

The GLRT uses all the data (primary and secondary) in the likelihood maximization under each hypothesis. The AMF test makes no use of the primary vector to estimate the covariance; therefore, poorer detection performance might be expected. In the following sections, the performance loss is shown to be small and that, in certain situations, the AMF test will actually outperform the GLRT.

4.2 CFAR Behavior

We now show that the density of the AMF test statistic does not depend on the true covariance matrix under H_0, and thus, it gives a constant false alarm rate test.

Let $\mathbf{u} = \mathbf{R}^{-\frac{1}{2}}\mathbf{d}$, and $\mathbf{y} = \mathbf{R}^{-\frac{1}{2}}\mathbf{z}$. Then the test can be written

$$\frac{|\mathbf{u}^\dagger \mathbf{R}^{\frac{1}{2}} \hat{\mathbf{R}}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{y}|^2}{\mathbf{u}^\dagger \mathbf{R}^{\frac{1}{2}} \hat{\mathbf{R}}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{u}} = \frac{|\mathbf{u}^\dagger \tilde{\mathbf{R}}^{-1} \mathbf{y}|^2}{\mathbf{u}^\dagger \tilde{\mathbf{R}}^{-1} \mathbf{u}} \underset{H_0}{\overset{H_1}{\gtrless}} \alpha \quad , \tag{73}$$

where $\tilde{\mathbf{R}} \equiv \mathbf{R}^{-\frac{1}{2}} \hat{\mathbf{R}} \mathbf{R}^{-\frac{1}{2}}$. $\tilde{\mathbf{R}}$ is subject to the complex Wishart distribution with parameters K, N, and $\mathbf{I}$, which will be denoted $CW(K, N; \mathbf{I})$ [5].

Now a unitary transform is defined that rotates the whitened signal vector into the first elementary vector.

$$d\mathbf{e} = \mathbf{U}^\dagger \mathbf{u}, \quad \mathbf{e} = [1, 0, \ldots, 0]^\dagger \quad . \tag{74}$$

The first column of $\mathbf{U}$ is the whitened signal vector $\mathbf{u}$, and the other $N - 1$ columns form an arbitrary orthonormal basis for the orthogonal complement of the subspace spanned by $\mathbf{u}$. The test then becomes

$$t = \frac{|d\mathbf{e}^\dagger \mathbf{S}^{-1} \mathbf{x}|^2}{d^2 \mathbf{e}^\dagger \mathbf{S}^{-1} \mathbf{e}} = \frac{|\mathbf{e}^\dagger \mathbf{S}^{-1} \mathbf{x}|^2}{\mathbf{e}^\dagger \mathbf{S}^{-1} \mathbf{e}} \underset{H_0}{\overset{H_1}{\gtrless}} \alpha \quad , \tag{75}$$

35

where $\mathbf{x} \equiv \mathbf{U}^{\dagger}\mathbf{y}$ and $\mathbf{S} \equiv \mathbf{U}^{\dagger}\bar{\mathbf{R}}\mathbf{U}$. Then $\mathbf{x}$ is distributed $N(\mathbf{0},\mathbf{I})$ under H_0, and $\mathbf{S}$ is distributed $CW(K,N;\mathbf{I})$. The actual covariance does not appear in this equation or in the underlying density functions, and thus, this is a CFAR test. This test is independent of both the scale and the structure of the true covariance, in contrast to the simple unknown level CFAR characteristic of many common CFAR detectors.

4.3 Generalization of the Signal Model

The equations to determine the performance of this detector will be derived for a general signal case where the signal may or may not be in alignment with the look direction. In our model the signal is assumed to lie along some general direction vector $\mathbf{p}$; hence, the signal is now normally distributed $N(\mathbf{0},\mathbf{R})$ on H_0 and $N(b\mathbf{p},\mathbf{R})$ on H_1. The steering vector of the array is assumed to be $\mathbf{q}$.

The direction vectors may be normalized so that

$$\mathbf{p}^{\dagger}\mathbf{p} = \mathbf{q}^{\dagger}\mathbf{q} = 1 \quad , \tag{76}$$

and the following definitions are made:

$$A_{\mathbf{q}}^2 \equiv (\mathbf{q}^{\dagger}\mathbf{R}^{-1}\mathbf{q}) \quad , \tag{77}$$

and

$$A_{\mathbf{p}}^2 \equiv (\mathbf{p}^{\dagger}\mathbf{R}^{-1}\mathbf{p}) \quad . \tag{78}$$

Summarizing Kelly [30], these terms may be used to describe the signal-to-noise ratio. That is, the maximum signal-to-noise ratio is

$$SNR_{\mathbf{qq}} = |b|^2 A_{\mathbf{q}} \quad , \tag{79}$$

which is attained when the signal lies along the axis for which the detector is steered. There is a signal-to-noise ratio loss when the signal does not lie in the steering direction. The signal-to-noise ratio which results when the array is steered in the direction corresponding to $\mathbf{q}$ is

$$SNR_{\mathbf{qp}} = |b|^2 \frac{|\mathbf{q}^{\dagger}\mathbf{R}^{-1}\mathbf{p}|^2}{(\mathbf{q}^{\dagger}\mathbf{R}^{-1}\mathbf{q})} \quad . \tag{80}$$

36

$\mathbf{q}^\dagger \mathbf{R}^{-1}\mathbf{p}$ can be interpreted as an inner product of $\mathbf{p}$ and $\mathbf{q}$, and the definition that

$$\cos\theta e^{j\phi} \equiv \frac{(\mathbf{q}^\dagger \mathbf{R}^{-1}\mathbf{p})}{A_\mathbf{p}A_\mathbf{q}} \tag{81}$$

can be made [21].

Then $\cos\theta$ may be used to relate the signal-to-noise ratio to the maximum signal-to-noise ratio as [3]

$$a \equiv SNR_\mathbf{qp} = SNR_\mathbf{pp}\cos^2\theta \qquad . \tag{82}$$

$SNR_\mathbf{qp}$ can be thought of as the signal-to-noise ratio in the subspace spanned by the adapted steering direction, and likewise,

$$c \equiv SNR_\mathbf{pp}\sin^2\theta \tag{83}$$

can be viewed as the signal-to-noise ratio in the orthogonal subspace.

4.4 Performance Evaluation

4.4.1 Derivation of Test Performance

The analysis of this detector is similar to Kelly's analysis [14] and uses the same notation. Appropriate whitening and unitary transforms as defined in Section 4.2 are performed to reformulate the AMF test in the statistically equivalent form

$$\Lambda = \frac{|\mathbf{e}^\dagger \mathbf{S}^{-1}\mathbf{z}|^2}{(\mathbf{e}^\dagger \mathbf{S}^{-1}\mathbf{e})} \underset{H_0}{\overset{H_1}{\gtrless}} \alpha \qquad . \tag{84}$$

In Equation (75) $\mathbf{z}$ has been redefined in this equation to be the whitened rotated primary data vector $\mathbf{x}$.

---

[3]This definition is equivalent to the definition of available signal-to-noise ratio a in Equation (44) when $\mathbf{p} = \mathbf{q}$.

The steps made to form this representation are identical to those used to show that the test statistic is independent of the underlying covariance matrix. Here, because of the generalization of the signal model, the actual mean direction vector may not have been transformed to the first elementary vector. With this in mind, z is now normally distributed $N(0, I)$ under H_0 and $N(bA_p f, I)$ under H_1. The transformed covariance estimate S has the complex Wishart distribution $CW(K, N; I)$, and the signal direction vector is given by

$$\mathbf{f} \equiv \frac{1}{A_{\mathbf{p}}} \mathbf{U}^\dagger \mathbf{R}^{-\frac{1}{2}} \mathbf{p} \quad . \tag{85}$$

U is the unitary matrix required to rotate the whitened direction vector to the first elementary vector $e \equiv [1, 0, \ldots, 0]^\dagger$. Following Kelly's method [14], the vector f is decomposed into components parallel and orthogonal to e, decomposing S as well:

$$\mathbf{P} = \mathbf{S}^{-1} = \begin{bmatrix} \mathbf{P}_{AA} & \mathbf{P}_{AB} \\ \mathbf{P}_{BA} & \mathbf{P}_{BB} \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{AA} & \mathbf{S}_{AB} \\ \mathbf{S}_{BA} & \mathbf{S}_{BB} \end{bmatrix}^{-1} \quad , \tag{86}$$

$$T \equiv (\mathbf{P}_{AA})^{-1} = \mathbf{S}_{AA} - \mathbf{S}_{AB}\mathbf{S}_{BB}^{-1}\mathbf{S}_{BA} \quad , \tag{87}$$

$$\mathbf{y} \equiv \mathbf{z}_A - \mathbf{S}_{AB}\mathbf{S}_{BB}^{-1}\mathbf{z}_B \equiv \frac{\mathbf{v}}{\sqrt{\rho}} \quad , \quad \text{and} \tag{88}$$

$$\rho \equiv (1 + \mathbf{z}_B^\dagger \mathbf{S}_{BB}^{-1}\mathbf{z}_B)^{-1} \quad . \tag{89}$$

The test statistic can now be simplified to the form

$$\frac{|\mathbf{e}^\dagger \mathbf{S}^{-1}\mathbf{z}|^2}{(\mathbf{e}^\dagger \mathbf{S}^{-1}\mathbf{e})} = \frac{|\mathbf{y}|^2}{T} = (1 + \mathbf{z}_B^\dagger \mathbf{S}_{BB}^{-1}\mathbf{z}_B)\frac{|\mathbf{v}|^2}{T} = \frac{|\mathbf{v}|^2}{\rho T} \quad , \tag{90}$$

and thus, the test may be expressed

$$|\mathbf{v}|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \alpha \rho T \quad , \tag{91}$$

where v is now normally distributed $N(0, 1)$ or $N(\sqrt{\rho}A_{\mathbf{p}}b\cos\theta, 1)$.

The GLRT is expressed in a similar form without the ρ factor in the threshold. T is an independent random variable which is distributed chi-squared (χ^2) with L complex degrees of freedom. Equation (91) has the form of a scalar CFAR test, in which the threshold is multiplied by the random loss factor ρ.

The density of the loss factor ρ has been derived [30], and it is given by the formula

$$f(\rho) = e^{-c\rho} \sum_{m=0}^{L+1} \binom{L+1}{m} \frac{(N+L-1)!}{(N+L-1+m)!} c^m f_\beta(\rho; L+1, N+m-1) \quad , \quad (92)$$

where c is given in Equation (83), and we have defined $L \equiv K + 1 - N$. The central beta density function is defined by

$$f_\beta(x; n, m) = \frac{(n+m-1)!}{(n-1)!(m-1)!} x^{n-1}(1-x)^{m-1} \quad . \quad (93)$$

An alternative form for $f(\rho)$ may be derived by expressing this function in terms of the confluent hypergeometric function and using Kummer's first transformation [31] to yield [30]

$$f(\rho) = e^{-c} \sum_{m=0}^{\infty} \frac{c^m}{m!} f_\beta(\rho; L+1, N+m-1) \quad . \quad (94)$$

4.4.2 Evaluation of the Probability of False Alarm

The probability of false alarm (PFA) for the AMF test is calculated when the signal mean is equal to 0; consequently, the orthogonal SNR term c is 0, and the density functions for ρ reduce to the central beta density function. The probability of false alarm will have the same form as that of the GLRT [15,14] except for the presence of the factor ρ in the threshold. As shown [21], the probability of false alarm for the GLRT is given by

$$PFA_{GLRT} = \frac{1}{(1+\alpha)^L} \quad , \quad (95)$$

where $\alpha = \gamma/(1-\gamma)$, and γ is the threshold term of Equation (63). To determine the false alarm probability for the AMF test, the term α can be replaced with $\rho\alpha$, and the expectation with respect to the loss factor ρ can be taken to yield

$$PFA_{AMF} = \int_0^1 PFA_{AMF}|\rho\, f(\rho)d\rho = \int_0^1 \frac{f_\beta(\rho; L+1, N-1)}{(1+\alpha\rho)^L} d\rho \quad . \quad (96)$$

This has been evaluated through numerical integration, and also by means of an expansion into an infinite series, integration term by term, and derivation of truncation bounds, with the same results. Iterative procedures based on bisection and Newton's method have been used to find α when a particular PFA is specified.

4.4.3 Evaluation of the Probability of Detection

The conditional probability of detection for the AMF test, given ρ, may be expressed in a finite sum expression as [32]

$$PD_{AMF}|\rho = 1 - \frac{1}{(1+\alpha\rho)^L} \sum_{m=1}^{L} \binom{L}{m} (\alpha\rho)^m G_m \left(\frac{a\rho}{1+\alpha\rho}\right) \quad , \tag{97}$$

where a is the SNR component defined in Equation (82) and where G_m is the incomplete Gamma function

$$G_m(y) \equiv e^{-y} \sum_{k=0}^{m-1} \frac{y^k}{k!} \quad . \tag{98}$$

Unlike the GLRT, the AMF test includes the loss factor in the threshold as well as in the mean. The expectation of the conditional probability of detection with respect to the random variable ρ must be taken to evaluate the unconditional probability of detection. The probability of detection can be written as

$$PD_{AMF} = 1 - \int_0^1 \frac{1}{(1+\alpha\rho)^L} \sum_{m=1}^{L} \binom{L}{m} (\alpha\rho)^m G_m \left(\frac{a\rho}{1+\alpha\rho}\right) f(\rho)d\rho \quad . \tag{99}$$

This equation has been computed through numerical integration using the finite sum form of the density function of Equation (92). Additionally, this equation has been evaluated through the use of the infinite series form of the density function by integrating term by term to express the probability of detection as a series expression containing two finite and three infinite series. Bounds for the three infinite series were obtained using methods similar to that of Shnidman [33]. The results using this method were then used to verify the results of the numerical integration.

In order to evaluate the probability of detection numerically, a single routine has been written to evaluate the probability of detection of a scalar CFAR detector [32]:

$$PD_{cfar}(\alpha, a, L) = 1 - \frac{1}{(1+\alpha)^L} \sum_{m=1}^{L} \binom{L}{m} \alpha^m G_m \left(\frac{a}{1+\alpha}\right) \quad . \tag{100}$$

The numerical integration technique of finding the unconditional probability of detection consists of repeatedly calling the routine with a and α replaced by $a\rho$ and $\alpha\rho$, weighting the result by $f_3(\rho)$, and summing. When this procedure is performed for the AMF test, the equation implemented is

$$PD_{AMF} = \Delta\rho \sum_{i=0}^{I-1} PD_{cfar}(i\Delta\rho\alpha, i\Delta\rho a, L) f_\rho(i\Delta\rho), \quad \Delta\rho = \frac{1}{I} \quad , \tag{101}$$

with a defined in Equation (82) as the SNR component parallel to the direction vector. I is chosen to yield a suitably small error by successively doubling the number of terms until the probability of detection varies less than some ϵ. It typically requires between 1024 and 8192 integration terms to achieve an error bound of 0.00001.

The corresponding equation for the GLRT is numerically integrated in the same manner, with the equation implemented being

$$PD_{GLRT} = \Delta\rho \sum_{i=0}^{I-1} PD_{cfar}(\alpha, i\Delta\rho a, L) f_\rho(i\Delta\rho) \quad . \tag{102}$$

If a in Equations (101) or (102) is now replaced by γa, where γ is a random loss, the probability of detection for the Swerling target fluctuation models [3] may be found. For these cases, γ is subject to a χ^2 distribution with the number of degrees of freedom dependent upon the Swerling model chosen. For the Swerling I model [3], there will be only one complex degree of freedom, and the density function for γ is

$$f_\gamma = e^{-\gamma} \quad . \tag{103}$$

If the expectation with respect to γ is taken on $G_m(\gamma y)$ of Equation (98), then

$$\int_0^\infty G_m(\gamma y) f_\gamma d\gamma = 1 - \left(\frac{y}{1+y}\right)^m \quad , \tag{104}$$

and the conditional probability of detection is then found (after some algebra) to be [34]

$$PD_{AMF}|\rho = \left(\frac{1+a\rho}{1+(\alpha+a)\rho}\right)^L \quad . \tag{105}$$

This expression can be numerically integrated with respect to the loss factor ρ.

Using the same procedure for the GLRT yields [34]

$$PD_{GLRT|\rho} = \left(\frac{1 + a\rho}{1 + \alpha + a\rho}\right)^L \qquad . \tag{106}$$

In these cases, no signal mismatch is assumed. For comparison, the probability of detection for the Swerling I known covariance matched filter is [32]

$$PD_{MF} = e^{-\alpha/(1+a)} \qquad . \tag{107}$$

In Figure 7, it can be seen that the AMF detector requires a slightly higher signal-to-noise ratio to achieve the same probability of detection as the unconstrained GLRT detector for lower signal-to-noise ratios, with a crossover for the probability of detection of 0.9. The known covariance detector performs better than either of the two adaptive approaches. These detectors may be utilized if the increase in signal-to-noise ratio required to achieve the performance of the known covariance detector is acceptable. When the quantity of data is limited, the unconstrained detectors will not exist since the unconstrained maximum-likelihood estimate of the covariance matrix requires that $K \geq N$. For $K = N$, tens of decibels more signal-to-noise ratio may be required to provide the same detection performance as the known covariance detector, as is shown in the first of the plots. For a probability of detection of 0.8, the GLRT and the AMF detectors require approximately 11 and 12 dB greater signal-to-noise ratio for $N = 4, K = 4$. This loss in performance is unacceptable in most applications and motivates us to use our knowledge of the structure of the covariance matrix to increase the adaptive detector performance.

The basic idea found in the derivation of the AMF detector will be utilized again when a constrained estimate of the covariance matrix is substituted into the known covariance matched-filter detector. In Chapter 9, the performance of detectors based on constrained and unconstrained covariance estimates will be compared.
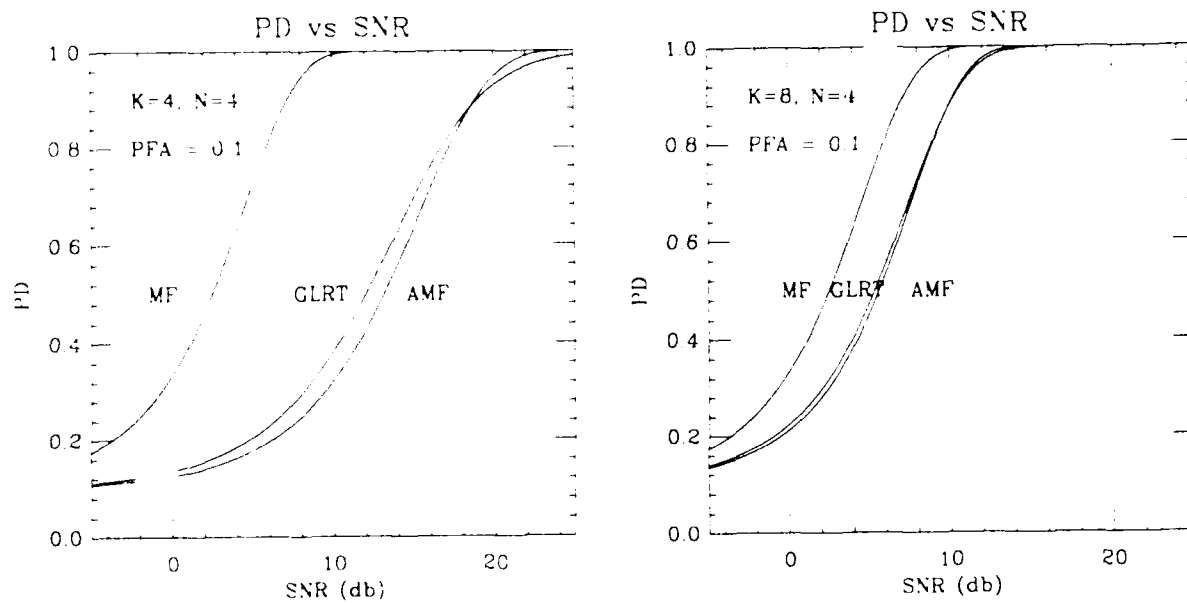
Figure 7. Probability of Detection versus SNR. Known covariance detector compared to unconstrained GLRT and AMF detectors.

5. COVARIANCE STRUCTURE

5.1 Introduction

The complete signal model derived in Chapter 2 is not commonly used in the derivation of adaptive signal-processing algorithms. Estimates of the covariance matrix, used to form the test statistic or weight vector, are typically not restricted to covariance matrices that are feasible with a particular array or noise environment. In subsequent chapters, the complete signal model will be used in the derivation of the adaptive beamforming and detection algorithms. In this chapter, covariance matrices that are consistent with the signal model are characterized, and methods of representing the matrices are presented. The parameters provided by the representations enable the use of the maximum-likelihood procedure to estimate covariance matrices that are constrained as a result of the data model.

The noise at the output of the sensor is assumed to be the sum of two independent terms. The first term is receiver noise, which is assumed to be independent from sensor to sensor. The contribution of this noise to the output covariance matrix is then a diagonal matrix and may be known. The effect of this matrix on the structure of the sensor output noise covariance matrix will be investigated later in this chapter.

The second noise term is due to the radiation received by the sensors. The responses of the sensors will not, in general, be mutually independent since the cross covariance between the received signals is assumed to satisfy the signal model found in Equation (32) and developed in Chapter 2. The contribution of this noise term to the output covariance matrix will lie within a class of matrices Ω_s. This class is characterized by an integral expression involving a non-negative spatial-temporal power-spectral-density function of Equation (32). Non-negativity of the spatial-temporal power-spectral-density function and linearity of the integral expression imply that the covariance matrix will exhibit a structure that is dependent upon the spatial locations of sensors, the spatial and temporal response of the sensors, and the transmission medium. The actual covariance matrix will be dependent upon the intensity and location of the spatially distributed noise sources.

The class of possible covariance matrices will be examined from geometric and algebraic viewpoints. This provides methods of representing the covariance matrices by the use of a finite number of terms and provides the base upon which adaptive signal-processing algorithms will be built.

5.2 Characterization of the Constraint Space

5.2.1 Constraint Space Definition

The covariance matrix has been shown to result from an integral expression involving a non-negative spatial-temporal power-spectral density along with functions that are defined by the

sensor responses and the geometry of Equation (32). The set of all matrices that satisfy the spatial constraint will be denoted by Ω_s:

$$\Omega_s = \quad \{\mathbf{R} : \mathbf{R} = \tfrac{1}{2\pi} \int_{\theta,\phi} \tilde{S}_{vv}(\omega_o, \theta, \phi) \mathbf{d}(\omega_o, \theta, \phi) \mathbf{d}^\dagger(\omega_o, \theta, \phi) \cos\theta d\theta d\phi,$$

$$\tilde{S}_{vv}(\omega_o, \theta, \phi) \geq 0\} \qquad . \tag{108}$$

We would now like to characterize the set of covariance matrices that are members of Ω_s.

Covariance matrices that are members of Ω_s belong to a larger vector space whose elements are general N by N Hermitian matrices. For Hermitian matrices $\mathbf{R}_1$ and $\mathbf{R}_2$ and real scalars α and β, $\alpha\mathbf{R}_1 + \beta\mathbf{R}_2$ is a Hermitian matrix, and this can be used to define a vector space.[4]

An inner product,

$$< \mathbf{R}_1, \mathbf{R}_2 > = \mathrm{tr}(\mathbf{R}_1 \mathbf{R}_2) \qquad , \tag{109}$$

can be defined on this space, and for completeness, the obvious norm is

$$\|\mathbf{R}\|^2 = \mathrm{tr}(\mathbf{R}\mathbf{R}) \qquad . \tag{110}$$

It will be convenient to write the elements of the covariance matrix $\mathbf{R}$ as an N^2 component real vector [35].

$$\mathbf{r} \equiv [r_1 \dots r_{N^2}]^T \qquad . \tag{111}$$

Let

$$r_i = R_{ii} \quad 1 \leq i \leq N \qquad , \tag{112}$$

and let

$$r_{N+1} + j r_{N+2} = \sqrt{2} R_{12} \qquad . \tag{113}$$

---

[4]Although I have used $\mathbf{R}$ to refer to the elements of this space, the members of this space include all Hermitian matrices, not just non-negative definite matrices.

By continuing this process along the upper minor diagonals of $\mathbf{R}$ in a generalized store-by-diagonal [36] ordering, then

$$\mathbf{r} \equiv [R_{11} \cdots R_{NN} \ \sqrt{2}\Re(R_{12}) \ \sqrt{2}\Im(R_{12}) \cdots \sqrt{2}\Re(R_{1N}) \ \sqrt{2}\Im(R_{1N})]^T \quad . \tag{114}$$

The precise ordering of the elements of $\mathbf{r}$ is unimportant to the results that utilize this notation. Because of Hermitian symmetry, $\mathbf{R}$ is fully specified by writing its upper triangle in this form.

With this definition of $\mathbf{r}$, there is an equivalence between the familiar vector inner product and the inner product given by Equation (109).

$$< \mathbf{R}_1, \mathbf{R}_2 > = \text{tr}(\mathbf{R}_1 \mathbf{R}_2) = \mathbf{r}_1^T \mathbf{r}_2 \quad . \tag{115}$$

5.2.2 Convex Cone

The space Ω_s of covariance matrices is a convex cone. This follows directly from the linearity of the integral and the positivity of the spatial-temporal power-spectral density. To see this, let S_1 and S_2 be valid distributions of spatial energy with covariance matrices $\mathbf{R}_1$ and $\mathbf{R}_2$, respectively, $\mathbf{R}_1, \mathbf{R}_2 \in \Omega_s$. Then, $S_3 = \alpha S_1 + \beta S_2$ is a valid spatial energy distribution for real $\alpha, \beta \geq 0$. The linearity of the covariance integral then implies that $\mathbf{R}_3 = \alpha \mathbf{R}_1 + \beta \mathbf{R}_2$ is a valid covariance matrix, so $\mathbf{R}_3 \in \Omega_s$.

The properties of convex sets can be used to parameterize the covariance matrices that are in our constraint space.

5.2.3 Dimension of Space

The space of possible covariance matrices, Ω_s, forms a subset within the set of Hermitian matrices. A characteristic of the subset, dependent upon the sensor responses and array geometry, is the *dimension* (L) of the space. The dimension L of Ω_s is the maximum number of elements of Ω_s that is linearly independent for real coefficients. That is, for $I > L$ there will exist a solution to

$$\sum_{i=1}^{I} \alpha_i \mathbf{R}_i = 0 \tag{116}$$

for real α_i not all zero.

We can readily see that the maximum dimension of Ω_s will be N^2. We will make the definition that the covariance matrices for an array exhibit *structure* if the dimension of Ω_s is less than N^2; Ω_s is a convex cone within a proper subspace of the N^2 real vector space. It will be assumed that the array is such that the covariance matrices will be structured.

47

5.2.4 Examples

The minimum dimension of Ω_s for unknown covariance matrices occurs when the noise environment is known except for a scale factor. The dimension of this space is 1. A CFAR detector based on this minimum constraint space is derived and analyzed in Appendix A. Closed-form expressions result for the test statistic and the equations that describe the performance of this detector. Because the model for the unknown covariance matrix in this case is the most restrictive constraint-cone possible, the performance of this test provides a convenient reference to which the performance of other constrained detectors can be compared.

It will be instructive to consider examples of what the dimension would be for other array geometries. The maximum dimension for sensors with identical spatial response is $N^2 - N + 1$. Covariance matrices for this array will have equal diagonal elements but otherwise may have considerable freedom for the values of the off-diagonal terms. The maximum dimension occurs for sensor spacings such that there are no repeated off-diagonal covariance "lags." This would be quite easy to achieve since sensor spacings are continuous quantities. Sensors can be spaced so that the three-dimensional differences in the array positions do not repeat. This set of array position differences is known as the co-array [37].

A linear array whose array spacings are a multiple of a common factor and whose covariance matrix will contain no repeated or missing off-diagonal lags is known as a *zero redundancy array*. A zero redundancy array cannot always be formed from N array elements. An array with the minimum number of repeated lags is known as a *minimum redundancy array*. Pillai [38] has tabulated the array spacings for zero and minimum redundancy arrays. Some of the simulation results shown in Chapter 9 will be for a minimum redundancy array.

For uniform linear arrays, we have seen in Section 2.7.2 that the underlying covariance matrix will exhibit the Toeplitz structure. Here, the dimension is $2N - 1$. A detection method based on restricting covariance matrices to this structure has been investigated by Fuhrmann [29].

If the eigenvectors of the covariance matrix are known, then there is a unitary transformation that can be applied to the data so that the covariance matrix is diagonal. This situation occurs when the spatial noise field is assumed to be periodic, with the period given by the array aperture for a uniform linear array. In this instance, the dimension of the space will be N. Fuhrmann [39] has investigated detection algorithms based on this constraint space as well.

5.3 Representations

In order to develop adaptive signal-processing algorithms that make use of the constraints induced by the signal model, methods of enforcing the constraints while performing the parameter estimation must be developed. The methods that are developed here are based on two approaches. The first approach is to form a representation for the entire space of possible covariance matrices. Approximating the integral equation that defines the constraint space by a finite sum is the method utilized to generate this representation for the constraint space.

The second approach is through the representation of a covariance matrix as a member of a convex set. The characterization of the constraint space as a convex set allows certain representations for covariance matrices that are members of the set.

Both of the approaches develop representations parameterizing the covariance matrices with a finite number of terms. The finite representations allow the practical implementation of covariance estimators with estimates constrained to be within Ω_s of Equation (108).

5.3.1 Hyperplane Characterizations

Before further discussion of the representations, the algebraic and geometric idea of a hyperplane will be introduced. A *hyperplane* $\mathcal{H}$ in a linear vector space Ω can be defined as [40,41][5]

$$\mathcal{H} \equiv \{\mathbf{R} : \text{tr}(\mathbf{RG}) = 0, \mathbf{G} \neq 0, \mathbf{G} = \mathbf{G}^{\dagger}; \mathbf{G}, \mathbf{R} \in \Omega\} \qquad . \qquad (117)$$

The hyperplane that results from this definition is a maximal proper subspace of Ω. This equation defines an algebraic subspace, which is orthogonal to $\mathbf{G}$. The space of possible constrained covariance matrices is a reduced dimensional subset of all possible covariance matrices and is then a subset of a hyperplane of the larger dimensional space.

Hyperplanes are of interest because the boundaries of cones formed from a finite number of terms are hyperplanes or intersections of hyperplanes. The inner product of any element on the hyperplane boundary of the constraint set with at least one of the elements of the vector space will be zero.

Additionally, there are some representations for covariance matrices that can be viewed as hyperplane representations.

5.3.2 Finite Sum Representation

The first representation that will be used to parameterize the constraint space is to approximate the covariance integral found in Equation (28) by an M-term finite sum[6]

$$\tilde{\Omega}_s = \{\mathbf{R} : \mathbf{R} = \sum_{m=1}^{M} S(\omega_o, \theta_m, \varphi_m)\mathbf{d}(\omega_o, \theta_m, \varphi_m)\mathbf{d}^{\dagger}(\omega_o, \theta_m, \varphi_m)\} \qquad , \qquad (118)$$

---

[5]More generally, $\text{tr}(\mathbf{RG}) = c$, an arbitrary constant. Our attention is placed on hyperplanes that pass through the origin.

[6]The differential area or volume is a positive quantity, which can be incorporated into the scaling of the direction vectors or the spatial spectrum.

where the number of terms M and the locations of the spatial-temporal spectral points (θ_m, ϕ_m) have been chosen to produce a suitable approximation to the integral of Equation (28). One such distribution of points is spatially uniform in (θ, ϕ). The effect of this is to approximate the convex cone of the constraint set by an inscribed polyhedral cone [42,37]. This representation will be used in Chapter 6.

The outer products of the direction vectors are basis functions in the sense that any covariance matrix interior to the polyhedral cone can be formed by the weighted sum of these basis functions. These functions are not, in general, independent and may not form an algebraic basis for the covariance matrices that are in the constraint set. The definition of the cone requires that the weights for the basis functions be non-negative; the algebraic basis does not make this restriction.

If the bases are fixed, then not all matrices that are members of the constraint class can be represented in this manner for finite M. If the covariance matrix contains the contribution from a single large interference source, then it may be that there is not a representation of it in terms of the discrete spectrum. For a finite number of spectral points, the set of covariance matrices that cannot be represented is non-empty. If the number of points in our approximation M is increased and an upper bound is placed on $\int S(\bar{e}) \, d\bar{e}$, then the continuity of $\mathbf{d}(\omega_o, \theta, \phi)$ allows the representation to get arbitrarily close[7] to any matrix within Ω_s. An analogy to this would be to approximate a disk by the interior of an inscribed polygon. As the number of sides on the polygon is increased, the area of the disk that is not contained within the polygon can be made arbitrarily small. Figure 8 illustrates this effect for the covariance matrix with two identical sensors spaced one-half wavelength apart. If the covariance matrices possible when $R_{11} = 1$ are examined, then R_{12} will be constrained by the requirement that the covariance matrix be non-negative definite. The possible locations of R_{12} will be constrained to lie within a disk such that $|R_{12}| \le 1$. The approximation to this constraint space for four terms in the finite sum is shown by the inscribed square. This example will be investigated in more detail in Section 5.5.

It will be assumed that a large number of terms will be required to parameterize covariance matrices that are members of Ω_s. Spatial receiving arrays often have high-powered discrete interference sources; the covariance matrices that result are poorly conditioned. If the spatial spectrum is known to be "smooth," then this information could be used to reduce the number of terms in the summation.

5.3.3 Carathéodory and Maximal Representations

The large number of terms used in the previous section to parameterize the constraint space is not necessary to represent a single covariance matrix that is a member of this constraint space. The spectral weights of the previous section are not generally unique due to the linear dependence of the

---

[7]The norm of the difference between any covariance matrix within Ω_s and the nearest covariance matrix within the polyhedral cone can be made arbitrarily small.

$\Im(R_{12})$

$\Re(R_{12})$
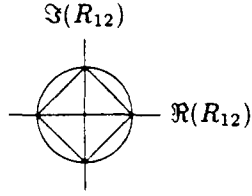
Figure 8. Constraint regions and approximations.

basis functions. This algebraic property of the basis functions can be used to find a representation of the covariance matrix requiring fewer terms. The number of terms that will be used in this representation is dependent upon the dimensionality of the constraint space. In the previous section any member of the constraint set could be represented arbitrarily closely by adjusting the spectral weights.[8] This will not be true with the representations introduced here. Rather than adjusting the weights of a large number of basis functions, the weights and the directions of arrival of a much smaller number of bases can be adjusted. This parameterization of the members of the constraint set can be used to estimate structured covariance matrices. The intent here is to estimate a structured covariance matrix with a reduced number of terms. To accomplish this the polyhedral cone formed by the finite basis must be allowed to "move" to enclose the estimate.

Theorem 5.1 *Carathéodory Representation [40,41,43,44] Every element* **R** *in an L-dimensional space* Ω *that is a finite convex combination of elements of* Ω *can be represented as a convex combination of* $L + 1$ *or fewer elements of* Ω.

Proof [44].
The proof of this theorem is instructive because it is one method of reducing the large number of terms, used in the finite sum, to a smaller number of terms. Using the vector representation for the elements of the vector space, form augmented vectors

$$\mathbf{r}'_i = \begin{bmatrix} 1 & \mathbf{r}_i^T \end{bmatrix}^T \qquad . \tag{119}$$

---

[8] As the number of bases approaches infinity.

Let $\mathbf{r} \in \Omega$ be represented as

$$\mathbf{r} = \sum_{i=1}^{M} \sigma_i \mathbf{r}_i, \quad \sum_{i=1}^{M} \sigma_i = 1, \quad \sigma_i > 0 \qquad . \tag{120}$$

If the number of terms M is less than $L + 2$, the proof is complete, and there is no need to proceed further. Now let

$$\mathbf{r}' = \sum_{i=1}^{M} \sigma_i \mathbf{r}_i' \qquad . \tag{121}$$

Since $M > L + 1$ then the $\mathbf{r}_i'$'s are not independent, and there exist μ_i's that are not all zero such that

$$\sum_{i=1}^{M} \mu_i \mathbf{r}_i' = 0 \qquad . \tag{122}$$

The first element of each vector is 1, so $\sum_{i=1}^{M} \mu_i = 0$, and at least one μ_i is positive. Let α be the largest number such that $\alpha \mu_i \leq \sigma_i$, $\forall i$. α is finite since at least one μ_i is positive. Now let $\dot{\sigma}_i = \sigma_i - \alpha \mu_i$. Then

$$\sum_{i=1}^{M} \dot{\sigma}_i \mathbf{r}_i = \sum_{i=1}^{M} \sigma_i \mathbf{r}_i - \alpha \sum_{i=1}^{M} \mu_i \mathbf{r}_i = \mathbf{r} \qquad , \tag{123}$$

and at least one $\dot{\sigma}_i = 0$. $\mathbf{r}$ has then been expressed in fewer elements. This argument can be repeated until $\mathbf{R}$ has been expressed as a positive linear combination of at most $L + 1$ elements of Ω. The method used here to reduce the number of terms is known as the *reduction theorem* [44].

This theorem is easily generalized to a finite positive combination of the elements with $\sum_{i=1}^{M} \sigma_i = c$, an arbitrary positive constant. The convexity of the sum was not utilized in the proof. Additionally, when $\mathbf{r}$ is formed from the elements of a convex cone, $\mathbf{r}_i$ may be scaled so that the condition $\sum_{i=1}^{M} \sigma_i = 1$ holds.

If one of the terms for all of the vector elements is positive and non-zero, then there is no need to augment the vectors. The method used to prove the Carathéodory representation can be applied directly to the vectors $\mathbf{r}_i$ rather than to $\mathbf{r}_i'$, with the result that the vector $\mathbf{r}$ can be represented using L elements rather than $L + 1$.

For sensor arrays where at least one sensor has a non-zero response throughout space, one of the diagonal entries of the dyads in Equation (118) will be positive for all terms in the sum.

A covariance matrix that is formed by the positive weighted sum of M elements of Ω_s can be represented as a combination of at most L elements of Ω_s.

Representing a member of a convex set in terms of L non-zero weighted bases is a *maximal representation* [44] of that member.

5.3.4 Hyperplane Representation

There is an additional Carathéodory representation theorem presented in Grenander and Szegö [45] that has been used to show that for Toeplitz matrices there is a representation in terms of $N-1$ "spectral" points plus a diagonal term [38,46]. Thus, any Toeplitz matrix can be represented as

$$\mathbf{R} = \sum_{i=1}^{N-1} \sigma_i \mathbf{d}_i \mathbf{d}_i^\dagger + \sigma_0 \mathbf{I} \quad , \tag{124}$$

where the $\mathbf{d}_i$ terms are Vandermonde vectors,[9] and $\sigma_i \geq 0$. The spectral points used in this representation are not necessarily the same as the spectral points introduced in the finite representation of the constraint space in Section 5.3.2.

This is not the same Carathéodory representation theorem as that presented in Section 5.3.3, but it can be related if the representation of the $\sigma_0 \mathbf{I}$ term is investigated. The term $\sigma_o \mathbf{I}$ can be formed by the sum of the outer products of N orthogonal Vandermonde vectors, the discrete Fourier transform vectors. Then $\mathbf{R}$ is the positive weighted sum of $2N - 1 = L$ direction vectors.

This representation can be interpreted for our problem as a hyperplane representation. There will be at least one vector in the L-dimensional vector space that is orthogonal to the convex cone generated by the matrices $\mathbf{I}$ and the $\mathbf{d}_i \mathbf{d}_i^\dagger$. The intersection of this hyperplane and the convex cone Ω_s contains the covariance matrix $\mathbf{R}$.

This representation would appear to be a suitable representation for covariance matrices that are members of the constraint space generated by a uniform linear array. Any matrix within the constraint space will have a representation in this form; however, the converse is not true. Matrices that can be represented in this form may not be members of the constraint space. The Vandermonde vectors that are used in this representation may not be realizable as direction vectors for a particular array geometry and operating frequency, and thus, the matrix could be outside of the constraint region. It is not clear from this representation whether a covariance matrix is in the interior or

---

[9]A Vandermonde vector is a (complex) N-vector such that $\mathbf{d}_i \equiv [\alpha_i^0 \; \alpha_i^1 \; \cdots \alpha_i^{N-1}]^\dagger$. In this case, $|\alpha_i| = c$. Uniform linear arrays have the property that the direction vectors are a scalar multiple of Vandermonde vectors.

exterior of the constraint region, and other methods would be needed to determine whether the matrix was realizable for a particular array geometry. This inadequacy of this representation will be illustrated in a later section.

5.4 Inclusion of Receiver Noise

We have been investigating the space of covariance matrices that is generated by a spatial distribution of energy. It can also be assumed that there is an additional additive noise term that is attributed to receiver noise. The noise covariance matrix attributed to this term will add to the output covariance matrix for spatially distributed noise. There will then be a constraint space for the output covariance matrix that is the cone of the spatially generated noise with the apex of the cone shifted away from the origin by the covariance matrix of the receiver noise. If the noise covariance matrix is unknown, then the space of possible output covariance matrices will be the constraint space for the spatially generated noise shifted over all possible shifts. Unless the noise covariance matrix is structured, then the resulting covariance matrix cannot be structured. Fortunately, the covariance matrices of the additive receiver noise are usually structured.

If the receiver noise is assumed to be independent but unknown, the dimension of the constraint space may be changed, resulting in additional terms that must be added to the basis functions in order to represent the constraint space or a covariance matrix. The basis functions that can be added to account for this are the outer products of the elementary vectors.[10] If the covariance matrix due to the spatially distributed noise sources can be represented as $\sum_{i=1}^{L} \sigma_i \mathbf{d}_i \mathbf{d}_i^\dagger$, the entire covariance matrix can be represented by

$$\mathbf{R} = \sum_{i=1}^{L} \sigma_i \mathbf{d}_i \mathbf{d}_i^\dagger + \sum_{i=L+1}^{L+N} \sigma_i \mathbf{e}_{L-i} \mathbf{e}_{L-i}^\dagger \qquad . \tag{125}$$

If the set of all shifts is within the constraint space (e.g., due to independent and identically distributed noise sources for uniform arrays), then the combined noise covariance matrix will retain the structure of the spatial covariance matrix, and the representations for the constraint space and covariance matrix will apply to the combined constraint space.

Often the receivers are fully characterized, and the receiver noise variances can be assumed known. This information can be utilized in the estimation procedure for the covariance matrices.

5.5 Discussions and Examples for Two Sensors

Examples of constraint spaces and the representation of some elements of these constraint spaces will now be shown. The examples are for 2 by 2 covariance matrices. The arrays that are

---

[10] An elementary vector $\mathbf{e}_i$ is zero except for element i, which is unity.

used to form the matrices are assumed to be narrowband arrays, and the array element responses are identical. The covariance matrices that result for these arrays are *Toeplitz* because the diagonal elements will be equal, and thus, the dimension of the constraint space is 3.

Figures will be shown to illustrate the constraint spaces. These figures are generated by noting the covariance matrices that result from impulsive spatial spectra or a discrete interference source. A single impulse in the spatial spectra will give rise to a covariance matrix that is on the boundary of the constraint space. This matrix is a weighted outer product of a direction vector with itself. In this case, the direction vector is simply a vector of phase shifts; i.e., $\mathbf{d} = [1\, e^{j\theta_e}]^T$. The covariance matrices that are generated by a spectral impulse are matrices of the form

$$\mathbf{R} = \begin{bmatrix} R_{11} & R_{11}e^{-j\theta_e} \\ R_{11}e^{j\theta_e} & R_{11} \end{bmatrix} \qquad . \qquad (126)$$

From Chapter 2, for a narrowband source at a spatial direction θ_s, weight σ, and wavelength λ and with the distance between sensors s, then

$$R_{12} = \frac{\sigma}{2\pi}e^{j2\pi s\cos(\theta_s)/\lambda} = R_{11}e^{j\theta_e} \qquad . \qquad (127)$$

Varying the direction of arrival of the source θ_s will vary θ_e over a range that is dependent upon the source wavelength, the spacing of the array elements and the possible variations in the direction of arrival.

The constraint space is then generated as a convex combination of these boundary elements. The covariance matrices resulting from a single impulse in the spatial spectra are singular since the determinate of the matrix is 0.

Two types of figures are used to illustrate the constraint space and the representations. The first figure is a perspective view of an open cone, which is used to illustrate the covariance matrices that are possible for the constraint spaces. This figure is plotted in the coordinate system $[R_{11} \ \Re(R_{12}) \ \Im(R_{12})]^T$. The second figure is a cross section of the cone for $R_{11} = R_{22} = 1$. The coordinate axis has been rotated by 90 degrees for clarity. The reference $\theta_e = 0$ is along the $\Re(R_{12})$ axis.

5.5.1 Covariance Matrices Possible

The first example will be for a narrowband receiver array where the two elements are spaced one-half of a wavelength apart. Both of the elements are assumed to have the same spatial response and gain. For this array, the set of possible covariance matrices is an open cone, as shown in Figure 9. This is a direct result of the integral that defines the space of covariance matrices. Substituting

the element separation into Equation (127), the exponent is $j\pi\cos(\theta_s)$. Varying θ_s over the domain $0 \le \theta_s \le \pi$ results in the range for the electrical angle $-\pi \le \theta_e \le \pi$.

$$\Re\{R_{12}\} \qquad\qquad \Re\{R_{12}\}$$

$$R_{11}, R_{22} \quad \Im\{R_{12}\}$$

$$\Im\{R_{12}\}$$
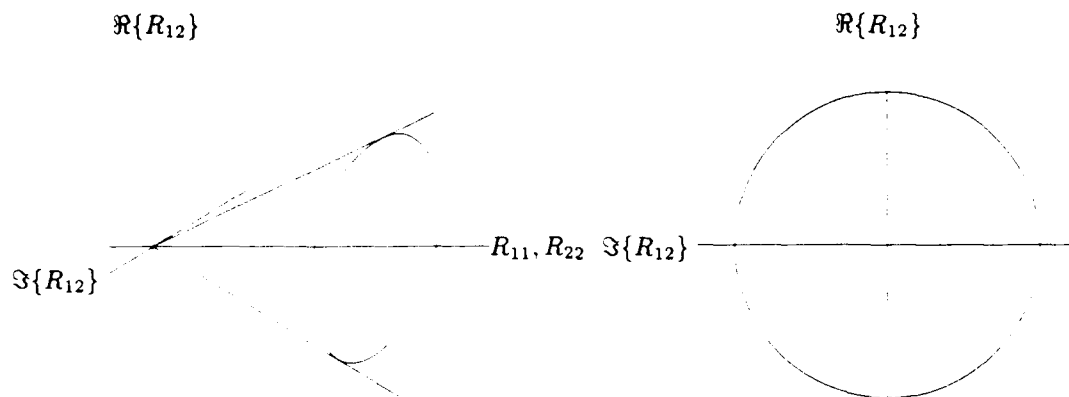
Figure 9.   Constraint space for one-half wavelength spacing.

The second example will be for a narrowband array where the array elements are placed three-eighths of a wavelength apart. The set of possible covariance matrices for this array is also an open cone but has a hyperplane for a portion of the boundary rather than having the continuous support of the previous example. This is shown in Figure 10. This is a result of the mapping from the physical angle of the source to the electrical angle. Proceeding as for the previous example, the exponent for the R_{12} term is $j\frac{3}{4}\pi\cos(\theta_s)$. Varying θ_s over the domain $0 \le \theta_s \le \pi$ results in the range for the electrical angle $-\frac{3}{4}\pi \le \theta_e \le \frac{3}{4}\pi$.

The range of θ_e here has been restricted by the spacing of the array elements. This restriction on θ_e could also be achieved if the domain of the angles of incidence of possible interferers is restricted.

5.5.2   Representations

The first representation is to approximate the constraint set by the use of the finite approx-imation to the covariance integral. The cone formed by the finite basis is shown in Figure 11 for four terms in the sum as well as for eight terms. The terms in the summation are spaced such that the covariance lags resulting from this spacing are uniformly spaced in electrical angle.

$\Re\{R_{12}\}$ $\Re\{R_{12}\}$

R_{11}, R_{22} $\Im\{R_{12}\}$

$\Im\{R_{12}\}$
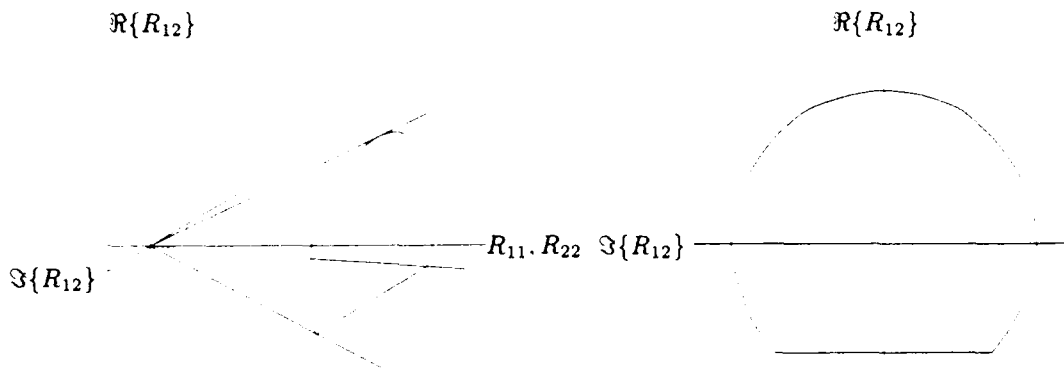
Figure 10.   Constraint space for three-eighth wavelength spacing.

It is assumed that a particular covariance matrix is within the constraint set formed by the finite approximation. The number of terms in the summation used to generate this matrix can be reduced using the reduction theorem. For this example, there are several polyhedral cones that can be formed containing the covariance matrix. This is shown in Figure 12 for two of the cones that could result. The intersections of the polyhedral cone with the boundary of the constraint set indicate the frequencies that are used to represent the covariance matrix. This example illustrates the non-uniqueness of the spatial spectrum. Both of these representations of the covariance matrix are maximal representations.

Figure 13 illustrates a hyperplane that intersects the constraint space and has the indicated covariance matrix as an element. There are many hyperplanes that have this property. The hyperplane shown is the hyperplane that is indicated by the Grenander and Szegö theorem of Carathéodory. This represents the covariance matrix as a weighted sum of the vectors labeled A and B in the three-dimensional projection of Figures 13 and 14 or as a convex combination of the points labeled A and B in the cross section. For the array with three-eighths wavelength spacing, the vector (point) B is not a member of the constraint set, and there is not a physical angle for a source that would generate this vector. The spectral point indicated is not valid for this array geometry.

$\Re\{R_{12}\}$ $\Re\{R_{12}\}$

$\Im\{R_{12}\}$ $R_{11}.R_{22}$ $\Im\{R_{12}\}$

$\Re\{R_{12}\}$ $\Re\{R_{12}\}$

$\Im\{R_{12}\}$ $R_{11}.R_{22}$ $\Im\{R_{12}\}$
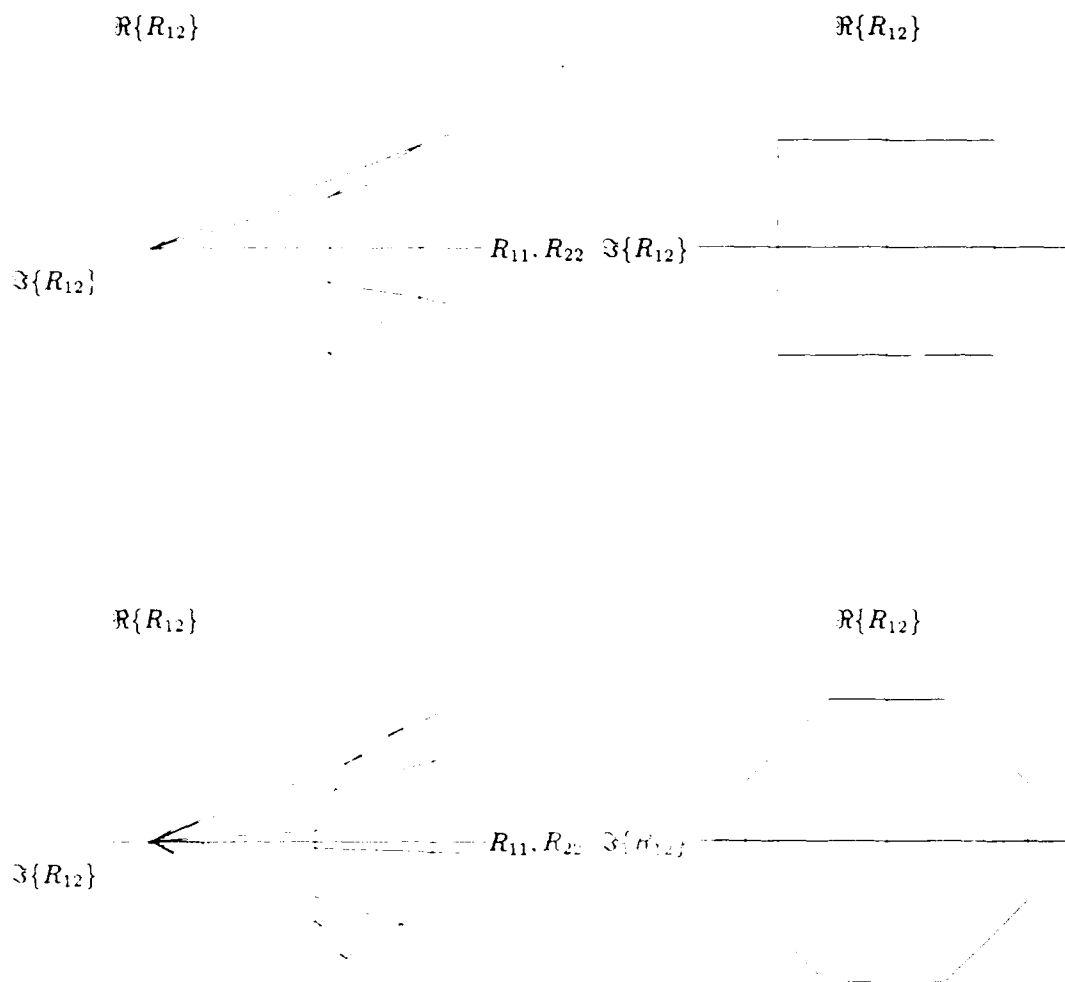
Figure 11.   Finite sum approximation and resulting constraint space.

58

$\Re\{R_{12}\}$ $\Re\{R_{12}\}$

R_{11}, R_{22} $\Im\{R_{12}\}$

$\Im\{R_{12}\}$

$\Re\{R_{12}\}$ $\Re\{R_{12}\}$

R_{11}, R_{22} $\Im\{R_{12}\}$

$\Im\{R_{12}\}$
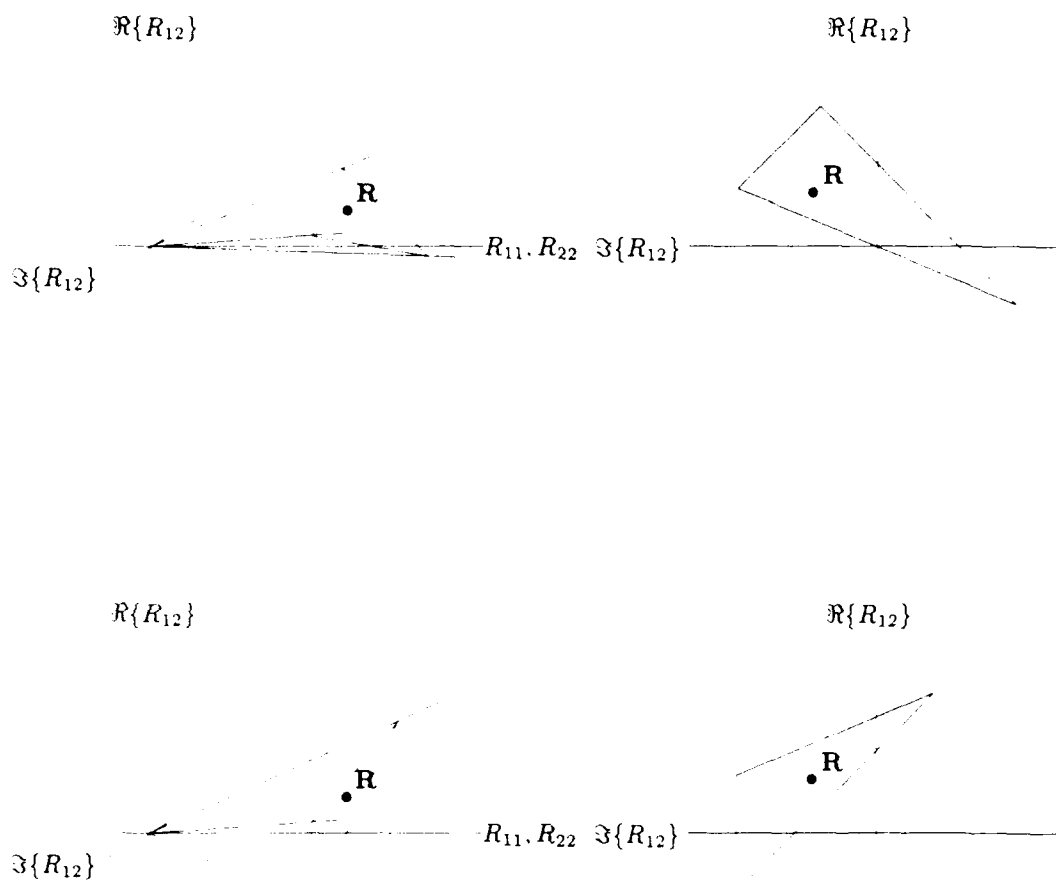
Figure 12.   Finite sum approximations after reduction.

59
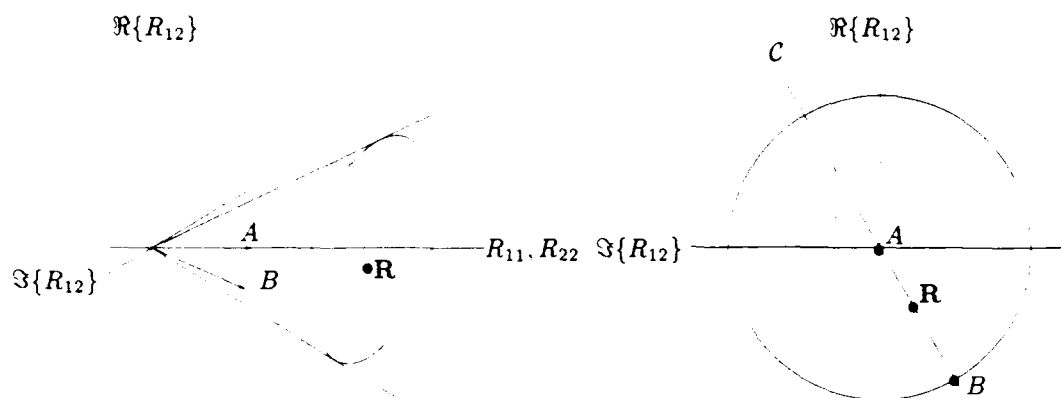
Figure 13.   Hyperplane representation.
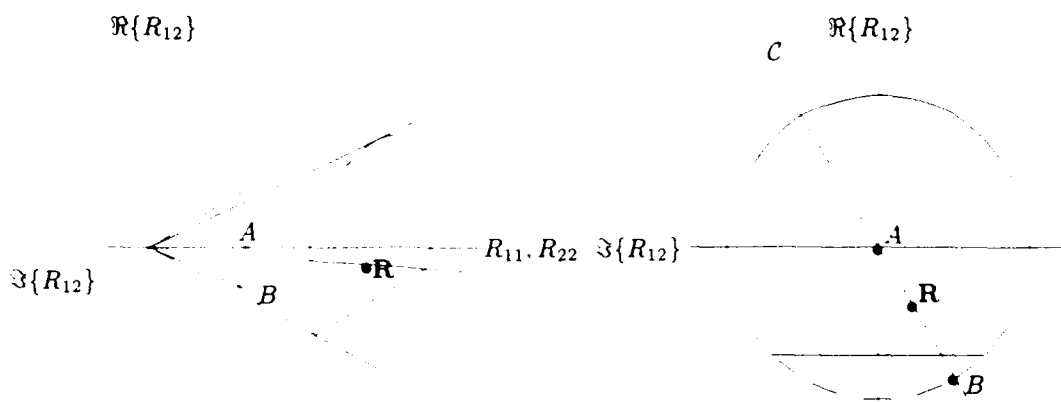
Figure 14.   Invalid Carathéodory representation.

5.6 Conclusion

Representations for the covariance matrices using a finite number or parameters have been introduced in this chapter. These representations and the properties of the constraint set will be used to form estimates of covariance matrices for use in adaptive signal-processing algorithms.

6. FIXED-BASIS ESTIMATOR

6.1 Introduction

The first approach that will be used to form joint estimates of the structured mean and covariance matrix is to approximate the integral expression that characterizes the constraint space using the finite sum representation discussed in the previous chapter. We will then derive an estimator for the mean and covariance matrix based on this spectral representation of the constraint space. The resulting estimator can be restricted so that only positive-definite estimates of the covariance will result. The finite sum approximation may introduce a bias for the covariance estimates since not all spatially constrained covariance matrices can be represented by the finite sum.

For simplicity the system is assumed to be narrowband. The covariance matrix will have the form

$$\mathbf{R}_{zz} = \frac{1}{2\pi} \int_{\theta,\phi} S(\omega_o, \theta, \phi) \mathbf{d}(\omega_o, \theta, \phi) \mathbf{d}^\dagger(\omega_o, \theta, \phi) \sin\theta \, d\theta d\phi + \mathrm{diag}(\eta_1^2 \ldots \eta_N^2) \qquad . \qquad (128)$$

The dependence on ω_o will be suppressed. Here, the terms that correspond to receiver noise are assumed unknown. Later, the modifications necessary when these terms are either known or assumed equal to each other will be discussed.

The data vectors are ordered such that the first G sample vectors are assumed to be non-zero mean, and the next K sample vectors are assumed to be zero mean. The sample vectors are assumed to be independent and are arranged in a data matrix $\mathbf{Z} = [\mathbf{z}_1 \cdots \mathbf{z}_P]$, $P = G + K$. The mean is of the form $b_k \mathbf{d}(\bar{\theta}, \bar{\phi})$, corresponding to an unknown amplitude b_k multiplying a known direction vector parameterized by the direction of arrival $\bar{\theta}, \bar{\phi}$. The multivariate density functions underlying all of these samples are assumed to share the same covariance matrix. The log-likelihood for the data is then given by

$$
\begin{aligned}
l \;=\; & -NP\log\pi - P\log\|\mathbf{R}\| \\
& - \sum_{k=1}^{G} (\mathbf{z}_k - b_k \mathbf{d}(\bar{\theta}, \bar{\phi}))^\dagger \mathbf{R}^{-1}(\mathbf{z}_k - b_k \mathbf{d}(\bar{\theta}, \bar{\phi})) - \sum_{k=G+1}^{P} \mathbf{z}_k^\dagger \mathbf{R}^{-1} \mathbf{z}_k \qquad . \qquad (129)
\end{aligned}
$$

We would like to find a maximum-likelihood estimator for the mean and the covariance under this model.

In this section the covariance estimates will be found using the M-term finite sum approximation for the covariance integral in Section 5.3.2.

63

$$\tilde{\mathbf{R}} = \sum_{m=1}^{M} S(\theta_m, \varphi_m) \mathbf{d}(\theta_m, \varphi_m) \mathbf{d}^\dagger(\theta_m, \varphi_m) + \text{diag}(\eta_1^2 \ldots \eta_N^2) = \mathbf{D\Sigma D}^\dagger \qquad . \qquad (130)$$

$\mathbf{D}$ is an $N\mathrm{x}(N+M)$ matrix formed from an $N\mathrm{x}M$ array of steering vectors with the identity matrix appended.

$$\mathbf{D} \equiv \left[\mathbf{d}(\theta_1, \phi_1) \cdots \mathbf{d}(\theta_M, \phi_M) \vdots \mathbf{I}_N \right] \qquad , \qquad (131)$$

and $\mathbf{\Sigma}$ is a diagonal matrix with non-negative diagonal entries.

6.2 Derivation of the EM Algorithm

A closed-form expression for the maximum-likelihood estimator for the covariance has not been found; however, the Expectation-Maximization (EM) algorithm [47] can be used to derive an iterative algorithm where each iteration will remain within the constraint region and each iteration will not decrease in likelihood.

The EM algorithm consists of the following two steps.

1. E-step: Given a complete data set $\mathbf{Y}$, that if known would uniquely determine the observed data, compute

$$E\{l_{cd}(\mathbf{Z}; \mathbf{\Theta}) | \mathbf{Y}, \mathbf{\Theta}^m\} \qquad . \qquad (132)$$

$l_{cd}(\mathbf{Z}; \mathbf{\Theta})$ here is the complete data log-likelihood, not the log-likelihood for the measured data.

2. M-step: Find

$$\mathbf{\Theta}^{m+1} = \underset{\mathbf{\Theta}}{\text{argmax}}\ E\{l_{cd}(\mathbf{Z}; \mathbf{\Theta}) | \mathbf{Y}; \mathbf{\Theta}^m\} \qquad . \qquad (133)$$

The EM algorithm can be shown to produce a sequence of estimates for which the likelihood is non-decreasing.

For the EM algorithm two data spaces must be defined. The complete data space is a hypothetical data space, and we will choose $M + N$-length vectors $\mathbf{y}$ arranged in a matrix $\mathbf{Y} \equiv [\mathbf{y}_1 \cdots \mathbf{y}_P]$ such that

$$\mathbf{Z} = \mathbf{DY} \qquad . \qquad (134)$$

The covariance of the columns of $\mathbf{Y}$ will be a diagonal matrix $\mathbf{\Sigma}$ with the first M diagonal terms being the discretized spectrum $\mathbf{S}$ and the next N diagonal terms corresponding to receiver noise. The incomplete data space is our observed data $\mathbf{Z}$. The matrix $\mathbf{D}$ will be ordered so that the

64

direction vector corresponding to the mean is the first column of the matrix $\mathbf{D}$ and will be denoted $\mathbf{d}_1$. The vectors $\mathbf{y}_k$ that correspond to the non-zero mean $\mathbf{z}_k$ will consist of a noise term plus an additive signal term that is a multiple of the first elementary vector

$$\mathbf{y}_k = b_k \mathbf{e}_1 + \mathbf{n}_k \quad , \tag{135}$$

with $\mathbf{d}_1 = \mathbf{D}\mathbf{e}_1$.

The complete data log-likelihood is

$$
\begin{aligned}
l_{cd} \;=\;& -(M+N)P\log\pi - P\log\|\mathbf{\Sigma}\| \\
& -\sum_{k=1}^{G}(\mathbf{y}_g - b_k\mathbf{e}_i)^{\dagger}\mathbf{\Sigma}^{-1}(\mathbf{y}_k - b_k\mathbf{e}_1) - \sum_{k=G+1}^{P}\mathbf{y}_k^{\dagger}\mathbf{\Sigma}^{-1}\mathbf{y}_k \quad .
\end{aligned}
\tag{136}
$$

If the diagonal entries of $\mathbf{\Sigma}$ are represented as $\sigma_1 \ldots \sigma_{M+N}$, the log-likelihood can then be written

$$
\begin{aligned}
l_{cd} \;=\;& -(M+N)P\log\pi - P\sum_{j=1}^{M+N}\log\sigma_j - \\
& \sum_{k=1}^{G}\frac{|Y_{1k} - b_k|^2}{\sigma_1} - \sum_{j=2}^{M+N}\sum_{k=1}^{G}\frac{|Y_{jk}|^2}{\sigma_j} - \sum_{j=1}^{M+N}\sum_{k=G-1}^{P}\frac{|Y_{jk}|^2}{\sigma_j} \quad .
\end{aligned}
\tag{137}
$$

The definition of $\mathbf{Y}$ has been used to avoid double subscripts since $Y_{jk} = (\mathbf{y}_k)_j$.

Define $\mathbf{B} \equiv [b_1 \cdots b_G]$. The expected value of the log-likelihood can be found (the E-step) to be

$$
\begin{aligned}
\mathrm{E}\{l_{cd}|\hat{\mathbf{\Sigma}}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\} \;=\; & \\
& -(M+N)P\log\pi - P\sum_{j=1}^{M+N}\log\sigma_j - \sum_{k=1}^{G}\frac{\mathrm{E}\{|Y_{1k} - b_k|^2|\hat{\mathbf{\Sigma}}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\}}{\sigma_1} - \\
& \sum_{j=2}^{M+N}\sum_{k=1}^{G}\frac{\mathrm{E}\{|Y_{jk}|^2|\hat{\mathbf{\Sigma}}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\}}{\sigma_j} - \sum_{j=1}^{M+N}\sum_{k=G+1}^{P}\frac{\mathrm{E}\{|Y_{jk}|^2|\hat{\mathbf{\Sigma}}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\}}{\sigma_j} \quad .
\end{aligned}
\tag{138}
$$

The third term in Equation (138) can be expanded as follows:

$$
\begin{aligned}
\mathrm{E}\{|Y_{1k} - b_k|^2|\hat{\mathbf{\Sigma}}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\} \;=\;& \mathrm{E}\{|Y_{1k}|^2|\hat{\mathbf{\Sigma}}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\} - 2\mathrm{Re}(b_k^* \mathrm{E}\{Y_{1k}|\hat{\mathbf{\Sigma}}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\}) + |b_k|^2 \\
\;=\;& \mathrm{E}\{|Y_{1k}|^2|\hat{\mathbf{\Sigma}}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\} \\
& -|\mathrm{E}\{Y_{1k}|\hat{\mathbf{\Sigma}}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\}|^2 + |\mathrm{E}\{Y_{1k}|\hat{\mathbf{\Sigma}}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\} - b_k|^2 \quad .
\end{aligned}
\tag{139}
$$

The conditional expectations can be written as

$$E\{\mathbf{y}_k|\hat{\boldsymbol{\Sigma}}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\} = \hat{b}_k^p \mathbf{e}_1 + \hat{\boldsymbol{\Sigma}}^p \mathbf{D}^\dagger \hat{\mathbf{R}}^{p^{-1}}(\mathbf{z}_k - \hat{b}_k^p \mathbf{d}_1) \tag{140}$$

and

$$E\{|Y_{jk}|^2|\boldsymbol{\Sigma}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\} = [\boldsymbol{\Sigma}^p - \hat{\boldsymbol{\Sigma}}^p \mathbf{D}^\dagger \hat{\mathbf{R}}^{p^{-1}} \mathbf{D}\hat{\boldsymbol{\Sigma}}^p]_{jj} + |E\{(Y_{jk}|\boldsymbol{\Sigma}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\}|^2 \tag{141}$$

for the data that have non-zero means, and

$$E\{|Y_{jk}|^2|\hat{\boldsymbol{\Sigma}}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\} = [\hat{\boldsymbol{\Sigma}}^p - \hat{\boldsymbol{\Sigma}}^p \mathbf{D}^\dagger \hat{\mathbf{R}}^{p^{-1}} \mathbf{D}\hat{\boldsymbol{\Sigma}}^p + \hat{\boldsymbol{\Sigma}}^p \mathbf{D}^\dagger \hat{\mathbf{R}}^{p^{-1}} \mathbf{z}_k \mathbf{z}_k^\dagger \hat{\mathbf{R}}^{p^{-1}} \mathbf{D}\hat{\boldsymbol{\Sigma}}^p]_{jj} \tag{142}$$

for the data that are zero mean.

Next, the arguments that maximize Equation (138) (the M-step) can be found. Maximizing this first with respect to the unknown b_k terms yields

$$\hat{b}_k^{p+1} = E\{Y_{1k}|\hat{\boldsymbol{\Sigma}}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\} = \hat{b}_k^p + [\hat{\boldsymbol{\Sigma}}^p \mathbf{D}^\dagger \hat{\mathbf{R}}^{p^{-1}}(\mathbf{z}_k - \hat{b}_k^p \mathbf{d}_1)]_1 \qquad . \tag{143}$$

Substituting this in, we next find the σ_j that maximize Equation (138).

$$\hat{\sigma}_j^{p+1} = \begin{cases} \dfrac{1}{P}\left[\displaystyle\sum_{k=1}^{K} E\{|Y_{jk}|^2|\hat{\boldsymbol{\Sigma}}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\} + \sum_{k=K+1}^{P} E\{|Y_{jk}|^2|\hat{\boldsymbol{\Sigma}}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\}\right] & j > 1 \\[2em] \dfrac{1}{P}\left[\displaystyle\sum_{k=1}^{K} E\{|Y_{1k}|^2|\hat{\boldsymbol{\Sigma}}^p, \hat{\mathbf{B}}, \mathbf{Z}\} \right. \\ \left. \qquad + \displaystyle\sum_{k=K+1}^{P} E\{|Y_{1k}|^2|\hat{\boldsymbol{\Sigma}}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\} - |E\{Y_{jk}|\hat{\boldsymbol{\Sigma}}^p, \hat{\mathbf{B}}^p, \mathbf{Z}\}|^2\right] & j = 1 . \end{cases} \tag{144}$$

The resulting iterations are

$$\hat{\boldsymbol{\Sigma}}_{jj}^{p+1} = \left[\hat{\boldsymbol{\Sigma}}^p + \hat{\boldsymbol{\Sigma}}^p \mathbf{D}^\dagger \left(\hat{\mathbf{R}}^{p^{-1}} \mathbf{S}_j^p \hat{\mathbf{R}}^{p^{-1}} - \hat{\mathbf{R}}^{p^{-1}}\right) \mathbf{D}\hat{\boldsymbol{\Sigma}}^p\right]_{jj} \qquad , \tag{145}$$

$$\hat{\mathbf{R}}^{p+1} = \mathbf{D}\hat{\boldsymbol{\Sigma}}^{p+1}\mathbf{D}^\dagger \qquad , \tag{146}$$

and

$$\hat{b}_k^{p+1} = \hat{b}_k^p + \left[\mathbf{\Sigma}^p \mathbf{D}^\dagger \ \hat{\mathbf{R}}^{p^{-1}} (\mathbf{z}_k - \hat{b}_k^p \mathbf{d}_1) \right]_1 \qquad , \tag{147}$$

where the sample covariance matrix is

$$\mathbf{S}_j^p = \begin{cases} \dfrac{1}{P} \left[\displaystyle\sum_{k=G+1}^{P} \mathbf{z}_k \mathbf{z}_k^\dagger \right] & j = 1 \\[2em] \dfrac{1}{P} \left[\displaystyle\sum_{k=1}^{G} (\mathbf{z}_k - \hat{b}_k^p \mathbf{d}_1)(\mathbf{z}_k - \hat{b}_k^p \mathbf{d}_1)^\dagger + \displaystyle\sum_{k=G+1}^{P} \mathbf{z}_k \mathbf{z}_k^\dagger \right] & j \neq 1 \end{cases} \tag{148}$$

If the true covariance were known, the EM iterations would not be needed since closed-form expressions for the unique maximizing b_k are well-known. This observation leads us to modify the EM iterations so that the likelihood with respect to the b_k is directly maximized in each iteration. The expressions for the b_k that will accomplish this are

$$\hat{b}_k^p = \frac{\mathbf{d}_1^\dagger \ \hat{\mathbf{R}}^{p^{-1}} \mathbf{z}_k}{\mathbf{d}_1^\dagger \ \hat{\mathbf{R}}^{p^{-1}} \mathbf{d}_1} \qquad . \tag{149}$$

Modifying the iteration rule for the b_k allows us to combine the iteration rules for all of the elements of $\mathbf{\Sigma}$ into a single iteration rule using the sample covariance matrix in Equation (148) for $j \neq 1$. The non-zero mean data will add to the sample covariance such that the iterations on $\mathbf{\Sigma}$ are unaffected by combining the iteration rules. This can be shown by evaluating products that would then appear in the iteration rule for $j = 1$.

$$\begin{aligned} \mathbf{d}_1^\dagger \hat{\mathbf{R}}^{p^{-1}} (\mathbf{z}_k - \hat{b}_k^p \mathbf{d}_1) &= \mathbf{d}_1^\dagger \hat{\mathbf{R}}^{p^{-1}} \mathbf{z}_k - \mathbf{d}_1^\dagger \hat{\mathbf{R}}^{p^{-1}} \mathbf{d}_1 \frac{\mathbf{d}_1^\dagger \ \hat{\mathbf{R}}^{p^{-1}} \mathbf{z}_k}{\mathbf{d}_1^\dagger \ \hat{\mathbf{R}}^{p^{-1}} \mathbf{d}_1} \\ &= 0 \qquad . \end{aligned} \tag{150}$$

It can then be seen that the iterations are unaffected by combining the rules. This also shows that the mean terms would not change for the next iteration of the formal EM algorithm. This term enters directly into the update equation for the mean. By making this modification to the formal EM algorithm, the direction vector corresponding to the mean need not be one of the columns of $\mathbf{D}$.

The algorithm that results from this modification is a generalization of the EM algorithm. It has the property that the likelihood produced by the sequence of estimates is non-decreasing, and for each iteration the increase in likelihood is lower bounded by the increase in likelihood of the formal EM iteration. This can be seen by examining the estimate of the mean produced by the formal EM iterations. If the likelihood is maximized with respect to the mean conditioned on the

67

current estimate of **R**, this would also be the estimate for the modified algorithm. If the likelihood is not maximized fully with respect to the mean, replacing the formal EM mean estimate with the modification will result in an increase in likelihood.

The generalized EM iterations are

$$\hat{\Sigma}_{jj}^{p-1} = \left[\hat{\Sigma}^p + \hat{\Sigma}^p \mathbf{D}^\dagger \left(\hat{\mathbf{R}}^{p^{-1}} \mathbf{S}^p \hat{\mathbf{R}}^{p^{-1}} - \hat{\mathbf{R}}^{p^{-1}}\right) \mathbf{D} \hat{\Sigma}^p\right]_{jj} \quad , \tag{151}$$

$$\hat{\mathbf{R}}^{p+1} = \mathbf{D} \hat{\Sigma}^{p+1} \mathbf{D}^\dagger \quad , \tag{152}$$

and

$$\hat{b}_k^p = \frac{\mathbf{d}_1^\dagger \hat{\mathbf{R}}^{-1} \mathbf{z}_k}{\mathbf{d}_1^\dagger \hat{\mathbf{R}}^{-1} \mathbf{d}_1} \quad , \tag{153}$$

where the sample covariance matrix is

$$\mathbf{S}^p = \frac{1}{P} \left[\sum_{k=1}^{G} (\mathbf{z}_k - \hat{b}_k^p \mathbf{d}_1)(\mathbf{z}_k - \hat{b}_k^p \mathbf{d}_1)^\dagger + \sum_{k=G+1}^{P} \mathbf{z}_k \mathbf{z}_k^\dagger\right] \quad . \tag{154}$$

If all of the samples are assumed to be zero mean, then this EM algorithm is similar to the algorithm of Miller and Snyder [26] and Moulin et al. [48] where it was proposed for Toeplitz constrained covariances.

The iterations for the spectral estimates take the form of a gradient ascent algorithm since the diagonal terms of

$$\mathbf{D}^\dagger \left(\hat{\mathbf{R}}^{p^{-1}} \mathbf{S}^p \hat{\mathbf{R}}^{p^{-1}} - \hat{\mathbf{R}}^{p^{-1}}\right) \mathbf{D} \tag{155}$$

are the gradients of the log-likelihood with respect to each of the elements of Σ. At stationary points of the iterative algorithms, the gradient will either be zero for the elements of Σ that are non-zero or will be negative or zero for elements of Σ that are zero. The estimates of the elements of Σ will never be negative as they are found in each iteration as conditional correlations. For these reasons the stationary points of the iterative algorithm coincide with local maxima of the likelihood.

If the receiver noise variances are assumed to be equal, then the algorithm is modified by replacing each of these terms by their average in each iteration. This is a straightforward modification that appears in the maximization step of the derivation. If the receiver noise variances (or

similarly the power level of the interference from any direction) are known, then these terms do not need to be estimated and can be replaced in each iteration by the known values.

6.3 Existence of Positive-Definite Solutions

In order to determine if a positive-definite solution exists to the problem characterized above, we will first assume that the receiver noise terms are known and equal to zero. Additionally, it will be assumed that the covariance matrix is estimated from data that are known to be zero mean.

The conditions necessary for the likelihood to be unbounded above were stated for a class of covariance matrices R satisfying certain restrictions [49].[11] Under these restrictions, the likelihood will be unbounded above if for a set of independent observations $\mathbf{Z} = \{\mathbf{z}_1 \cdots \mathbf{z}_K\}$, a singular matrix $\mathbf{R}_0$ exists such that $\mathbf{z}_k$ is in the range space of $\mathbf{R}_0$ for $k = 1 \cdots K$.

Our problem is parameterized such that the restrictions on the class of covariance matrices R are satisfied. The likelihood will be bounded above with probability 1 since the requirement that a singular matrix exists with the data in the range space is satisfied with probability 0, even for $K = 1$. There is a finite number of reduced dimensional subspaces corresponding to singular covariance matrices, and the probability that any $\mathbf{z}_k$ lies in one of these subspaces is 0.

Our attention will now turn to the case where all of the data are assumed to contain means that must be estimated. A maximum-likelihood estimate for this data set may not exist. If there exists a maximal proper subspace whose $N - 1$ basis vectors in $\mathbf{D}$ are orthogonal to $\mathbf{d}_1$, then the likelihood is unbounded above, and the maximum likelihood estimate of $\mathbf{R}$ does not exist. The existence of this subspace is a deterministic problem. and it sets conditions on the vectors that are the columns of $\mathbf{D}$.

Non-existence of the maximum-likelihood estimate can be shown by examining the likelihood as certain of the parameters of the likelihood are allowed to become arbitrarily small. Let the matrix $\mathbf{C}$ be the matrix formed from the weighted sum of outer products of the orthogonal subspace basis. $\mathbf{d}_1$ will be an eigenvector of $\mathbf{C}$ with associated eigenvalue 0. Let the other eigenvalues of $\mathbf{C}$ be denoted by $\delta_2 \ldots \delta_N > 0$. Now form a covariance matrix

$$\mathbf{R} = \mathbf{C} + \delta_1 \mathbf{d}_1 \mathbf{d}_1^\dagger \tag{156}$$

with inverse

$$\mathbf{R}^{-1} = \mathbf{C}^{-*} + \frac{1}{\delta_1} \mathbf{d}_1 \mathbf{d}_1^\dagger \quad . \tag{157}$$

---

[11] This paper was written with application to real symmetric matrices. It is not difficult to show that the conditions are true for complex matrices as well.

$\mathbf{C}^{-*}$ is the Moore-Penrose pseudoinverse of $\mathbf{C}$ [36]. $\mathbf{d}_1$ is an eigenvector of $\mathbf{R}$ and $\mathbf{R}^{-1}$.

Now substitute this covariance matrix into the log-likelihood functional for a single sample and investigate the effect of δ_1 becoming arbitrarily small. For a single sample, the log-likelihood with constant terms removed is

$$l = -\log \|\mathbf{R}\| - (\mathbf{z} - b\mathbf{d}_1)^\dagger \mathbf{R}^{-1}(\mathbf{z} - b\mathbf{d}_1) \qquad . \tag{158}$$

Substituting the maximum-likelihood estimate $\hat{b} = \dfrac{\mathbf{d}_1^\dagger \mathbf{R}^{-1}\mathbf{z}}{\mathbf{d}_1^\dagger \mathbf{R}^{-1}\mathbf{d}_1}$ and the matrix $\mathbf{R}$ and simplifying, the result is

$$l = -\sum_{i=1}^{N} \log \delta_i - \mathbf{z}^\dagger \mathbf{C}^{-*}\mathbf{z} \qquad . \tag{159}$$

As δ_1 approaches zero, the log-likelihood l approaches infinity; thus, the likelihood is unbounded above and a maximum-likelihood estimate will not exist. This will hold true for an arbitrarily large number of data samples provided a mean is to be estimated for each sample.

In the discretization of the integral equation for the covariance matrix, it was assumed that the discretization would adequately represent the resulting covariance matrix. In order to form this representation, a large number of spectral points and direction vectors may be needed. The existence of the condition allowing the likelihood to be unbounded can be determined by examining whether $N-1$ columns of $\mathbf{D}$ are orthogonal to $\mathbf{d}_1$. Restricting the basis vectors in this manner is not appealing; rather, the data model should be refined. It is known that the receiver noise variances are non-zero and can often be estimated accurately. Including the receiver noise variance in the covariance model will lower bound the minimum eigenvalue of the covariance matrix preventing a singular covariance estimate. The likelihood will then be bounded above.

6.4 Asymptotic Properties

The asymptotic behavior of the estimates will indicate an upper bound on the performance of any beamformer or detector that uses the estimates. We are interested in the behavior for two asymptotes: for the amount of data increasing and for the number of samples in the discrete approximation increasing.

First, the behavior of the covariance estimates when the number of zero mean data approaches infinity will be examined. The sample covariance matrix will converge to the true covariance matrix $\mathbf{R}_{true} \in \Omega_s$ with probability 1. If this matrix can be represented by the discretized model, then this will be the maximum-likelihood estimate, and this is a stable point of the EM algorithm. The estimator is then asymptotically unbiased.

If the true covariance is near singular, then there may not be a representation of the true covariance in terms of the discretized spectrum, and an asymptotically biased estimator results. All implementations of this estimator must be formed from a finite number of spectrum samples and will not represent all possible covariance matrices. The asymptotic bias of the estimator will result in a loss in likelihood compared to the maximum, impacting detectors that are based on likelihood. It would also be expected that the loss in signal-to-noise ratio for beamformers that use the covariance estimate would approach zero as the amount of data becomes large. Because the true covariance cannot be represented, the optimal weight vector will not, in general, be calculated, and a loss in signal-to-noise ratio will occur.

6.5 Brief Simulation Results

The performance gain of an adaptive beamformer based on this constrained estimator will be used to show that adaptive signal processing using constrained covariance estimates is viable. The loss in signal-to-noise ratio for adaptive beamformers using constrained and unconstrained estimates of the covariance matrix will be compared. This loss factor

$$\rho = \frac{|\mathbf{d}^\dagger \hat{\mathbf{R}}^{-1} \mathbf{d}|^2}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d} \mathbf{d}^\dagger \hat{\mathbf{R}}^{-1} \mathbf{R} \hat{\mathbf{R}}^{-1} \mathbf{d}} \tag{160}$$

is a random variable, and the histogram of the loss factor for constrained covariance estimates will be compared with the analytic expression for the density function when unconstrained estimates are used. The histograms are found using Monte Carlo simulation.

First, the performance is shown when the true covariance is within the polyhedral cone of the finite basis functions. The array simulated here is a 4-element minimum redundancy array with $K = N$. This array and the simulations will be discussed further in Chapter 9. Figure 15 illustrates the performance that can be obtained using structured estimates compared to the unconstrained covariance estimate. The line labeled ML is the density function of the signal-to-noise ratio loss factor when the unconstrained estimate of the covariance is used in an adaptive beamformer. The line labeled CML is an estimate of the loss factor density function when an adaptive beamformer is based on the constrained covariance estimate. This is the histogram for 200 realizations. The increase in the mean of the loss factor indicates that a dramatic increase in the performance of adaptive signal processors could be expected when using the constrained estimates.

In Figure 16 the effects of an insufficient number of terms in the finite sum are shown. The true covariance here is outside of the assumed polyhedral cone. The density function shows a high loss in signal-to-noise ratio when compared to the constrained covariance estimate with an adequate number of terms in the finite sum. A beamformer based on this covariance estimate would also show a loss when compared to the beamformer using an unconstrained covariance estimate.

In the next chapter, an alternative parameterization of the covariance will be introduced. This parameterization is based on the representation theorems introduced in Chapter 5 and will not
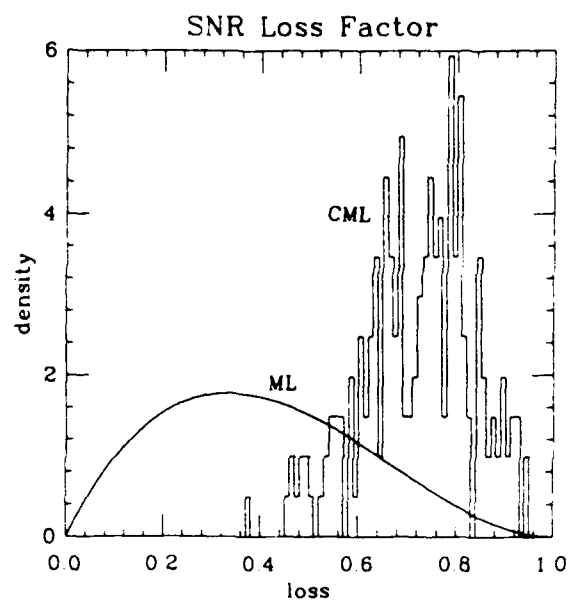
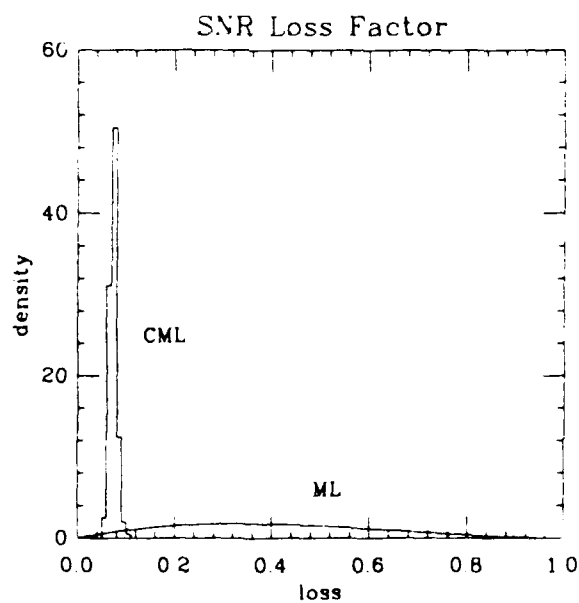*Figure 15. Signal-to-noise ratio loss factor density. Adequate number of terms.*

Figure 16. Signal-to-noise ratio loss fac or density. Inadequate number of terms.

require the large number of terms that may be required to approximate adequately the covariance integral.

7. VARIABLE BASIS ESTIMATOR

The reduction theorem and the Carathéodory representation theorem discussed in Chapter 5 state that any member of an L-dimensional convex set can be represented as the convex combination of $L + 1$ elements of that set. The large number of spectral weights used in the previous chapter to induce the constraint is mathematically unnecessary and, because of the large number of calculations needed to calculate the weights, is undesirable from an engineering perspective. In this chapter, a method is proposed that can be used to form joint estimates of the structured means and covariance matrices by using the maximal representation. The proposed method is capable of representing all covariance matrices satisfying the spatial constraint. The fixed-basis finite representation used in Chapter 6 is unable to accomplish this. The reduction in the number of spectral terms that must be estimated reduces the per-iteration computational demand and increases the rate of convergence of the resulting covariance estimates.

The Carathéodory representation theorem can be viewed as stating that there is an inscribed polyhedral cone formed from $L + 1$ bases containing an estimated covariance matrix. This cone cannot be formed a priori because the covariance matrix may take on values anywhere within the constraint region. For each estimated covariance matrix, a polyhedral cone must be found that contains the estimate in the interior. To determine this cone, the direction vectors used to form the basis functions must be allowed to change by varying the directions of arrival. This is very similar to the stochastic direction of arrival problem where it is assumed that the spatial source contribution to the covariance estimate lies on a boundary of the constraint space. The angles parameterizing the basis functions for this hyperplane are estimated to determine the estimated directions of arrival.

It would be desirable to use current direction-finding algorithms if they could be extended to fit our requirements. Unfortunately, these methods suffer from shortcomings that limit their usefulness for our purposes. Eigen-structure or subspace methods, such as multiple signal classification (MUSIC), require that the number of interference sources be limited to $N - 1$ sources and full rank sample covariance matrices. Previous stochastic direction-finding algorithms based on maximizing likelihood are known to perform poorly when there are many interference sources and the sources are spatially closer together than one-quarter of a beamwidth [50]. Additionally, the stochastic direction-finding algorithms require a computationally intensive search over the possible directions of arrival of the interference sources for each iteration.

A new procedure based on the estimator of Chapter 6 but which does not fix the parameters of the bases will be introduced. This estimation procedure allows the estimation of any covariance matrix that is within the constraint space. The likelihood is non-decreasing for the sequence of estimates provided by the algorithm, and stable points of the resulting algorithm satisfy the necessary conditions required for a maximum-likelihood estimate.

7.1 Preliminaries

In this section we will briefly digress in order to understand some items that will be used in the proposed estimation algorithms. Three topics of interest along with the Carathéodory theorem are used to justify the proposed algorithm. The first topic is how the likelihood surface will affect the constrained covariance estimates. The second topic is a discussion of how the discretization used in Chapter 6 affects convergence of the estimate for certain estimated covariance matrices. A maximum-likelihood spectral estimator is then discussed.

7.1.1 Likelihood Surface

The likelihood is unimodal over all covariance matrices for a non-singular sample covariance matrix and has only a single local maxima. If the estimate moves in any direction away from the unconstrained ML estimate, the likelihood is a monotonically decreasing function. Forming covariance estimates $\mathbf{R}_a = \mathbf{S} + \mathbf{C}$ and $\mathbf{R}_b = \mathbf{S} + \alpha\mathbf{C}$ with $\mathbf{R}_a, \mathbf{R}_b > 0$ and with $\mathbf{C} = \mathbf{C}^\dagger$ then, for $\alpha > 1$, $l(\mathbf{R}_a) \geq l(\mathbf{R}_b)$ with equality only if $\mathbf{C} = \mathbf{0}$. $\mathbf{S}$ is the sample covariance matrix that provides the maximum in the unconstrained case.

The unimodality of the likelihood is shown in the following proof. $\mathbf{S}$ and $\mathbf{C}$ are similar to diagonal matrices; therefore, there exists a transformation $\mathbf{A}$ that will simultaneously diagonalize $\mathbf{S}$ and $\mathbf{C}$ [28]. Define the diagonal matrices $\tilde{\mathbf{S}} = \mathbf{A}^{-1}\mathbf{S}\mathbf{A}$ and $\tilde{\mathbf{C}} = \mathbf{A}^{-1}\mathbf{C}\mathbf{A}$ with diagonal elements $\tilde{s}_1 \ldots \tilde{s}_N$ and $\tilde{c}_1 \ldots \tilde{c}_N$. This transformation is linear and will simultaneously diagonalize $\mathbf{R}_a$ and $\mathbf{R}_b$. Using these definitions and neglecting constant terms, the log-likelihoods are proportional to

$$l(\mathbf{R}_a) = -\sum_{i=1}^{N}\left[\log(\tilde{s}_i + \tilde{c}_i) + \frac{\tilde{s}_i}{\tilde{s}_i + \tilde{c}_i}\right] \qquad . \tag{161}$$

and

$$l(\mathbf{R}_b) = -\sum_{i=1}^{N}\left[\log(\tilde{s}_i + \alpha\tilde{c}_i) + \frac{\tilde{s}_i}{\tilde{s}_i + \alpha\tilde{c}_i}\right] \qquad . \tag{162}$$

The difference in log-likelihood between these estimates is

$$l(\mathbf{R}_a) - l(\mathbf{R}_b) = -\sum_{i=1}^{N}\left[\log\left(\frac{\tilde{s}_i + \tilde{c}_i}{\tilde{s}_i + \alpha\tilde{c}_i}\right) + \frac{\tilde{s}_i}{\tilde{s}_i + \tilde{c}_i} - \frac{\tilde{s}_i}{\tilde{s}_i + \alpha\tilde{c}_i}\right] \qquad . \tag{163}$$

Applying the inequality $\log x \leq x - 1$, then

76

$$l(\mathbf{R}_a) - l(\mathbf{R}_b) \geq \sum_{i=1}^{N} \left[1 - \frac{\tilde{s}_i + \tilde{c}_i}{\tilde{s}_i + \alpha\tilde{c}_i} - \frac{\tilde{s}_i}{\tilde{s}_i + \tilde{c}_i} + \frac{\tilde{s}_i}{\tilde{s}_i + \alpha\tilde{c}_i} \right] \quad , \tag{164}$$

with equality if and only if $\alpha\tilde{c}_i = \tilde{c}_i$. Simplifying this expression yields

$$l(\mathbf{R}_a) - l(\mathbf{R}_b) \geq \sum_{i=1}^{N} \frac{\tilde{c}_i^2(\alpha - 1)}{(\tilde{s}_i + \tilde{c}_i)(\tilde{s}_i + \alpha\tilde{c}_i)} \quad . \tag{165}$$

All of the terms that enter into this sum are non-negative for $\alpha \geq 1$, showing the unimodality of the likelihood surface. For $\alpha > 1$, the equality in Equation (165) holds only if all $\tilde{c}_i$ are zero corresponding to $\mathbf{C} = \mathbf{0}$. A unimodal surface such as this will have only a single extremum.

Unimodality of the likelihood surface for unconstrained maximum-likelihood estimates does not indicate the shape of the likelihood surface over more restricted covariance spaces. There may be ridges in the unconstrained likelihood surface even though it is unimodal. If a constraint set passes through several of these ridges, then local maxima will exist and the likelihood surface will be poorly behaved.

The likelihood is continuous for positive-definite covariance matrices. The determinant, matrix inverse, and the trace functions are all continuous for positive-definite arguments.

7.1.2   Constrained Covariance Estimates and Finite Bases

If the sample covariance is within the constraint space (such as occurs for $K \to \infty$), then the maximum-likelihood covariance estimate will be the sample covariance matrix. Any movement within the constraint space away from the sample covariance will result in a corresponding loss in likelihood. If the sample covariance matrix is not within the polyhedral cone, then the likelihood will not reach the maximum for estimates within the constraint space. When a constrained maximum-likelihood estimate is outside of the polyhedral cone formed by the finite basis, by continuity of the likelihood and unimodality over all matrices, the finite basis constrained estimate will lie on the boundary of the polyhedral cone. This can be proven by contradiction. Assume that there is a maximizer interior to the polyhedral cone. By the argument of Section 7.1.1, the likelihood will be increasing along the line from the interior point to the sample covariance exterior to the cone; therefore, the interior point cannot be a maximizer. This boundary is a hyperplane of dimension at most $L - 1$.

If an estimate lies on a hyperplane boundary, then only those basis elements that are in the subspace defined by the hyperplane may have non-zero weights. If any set of $L - 1$ bases are linearly independent, then at most $L - 1$ spectral weights will be non-zero. This condition can be easily tested; however, it is not altogether useful for our purpose.

The EM estimator performs poorly when the spectral weights are approaching zero, that is, when the covariance estimate is approaching a hyperplane boundary. Examining the update equation for a spectral weight

$$\sigma_j^{p+1} = \sigma_j^p + \sigma_j^p \mathbf{d}_j^\dagger \left(\hat{\mathbf{R}}^{p^{-1}} \mathbf{S}_j^p \hat{\mathbf{R}}^{p^{-1}} - \hat{\mathbf{R}}^{p^{-1}} \right) \mathbf{d}_j \sigma_j^p \quad , \tag{166}$$

it is seen that this can be written as

$$\sigma_j^{p+1} = \sigma_j^p (1 + \sigma_j^p \Delta) \quad . \tag{167}$$

The term Δ here is the gradient of the likelihood with respect to σ_j. The fractional change in the value of the spectral estimate will be given by $\sigma_j^p \Delta$. As the spectral estimate decreases, then the estimate will change by a smaller and smaller fraction of its value for each iteration. The value of Δ will be decreasing as well when the estimate moves closer to a maximum-likelihood estimate. For these reasons, the spectral estimates will approach zero very slowly. For finite word implementations roundoff error in calculating the update will result in a spectral estimate that does not change when $\sigma_j^p \Delta$ is smaller than the machine precision.

We would like for any estimate within the constraint region to be represented within some finite approximation to the constraint space. When the estimate is moving outside of the polyhedral cone, as evidenced by the solution on or approaching the hyperplane boundary, then the cone can be re-formed by adding additional bases in that direction.

7.1.3 A Maximum-Likelihood Spectral Estimator

A non-iterative solution for joint maximum-likelihood estimation of multiple independent spectral lines has not been found. However, the maximum-likelihood estimate of a single spectral line can be found conditioned on a previous estimate of the covariance matrix and any non-zero means. This problem can be stated as follows: Given a sample covariance matrix $\mathbf{S}$ formed from P mutually independent random variables

$$\mathbf{S} = \frac{1}{P} \left[\sum_{k=1}^G (\mathbf{z}_k - b_k \mathbf{d})(\mathbf{z}_k - b_k \mathbf{d})^\dagger + \sum_{k=G+1}^P \mathbf{z}_k \mathbf{z}_k^\dagger \right] \quad , \tag{168}$$

estimate a covariance matrix of the form $\mathbf{R} + \sigma \mathbf{d} \mathbf{d}^\dagger$. $\mathbf{R}$ and $\mathbf{d}$ are assumed to be known, and σ is unknown but positive. The estimate of σ is found by a maximum-likelihood estimator. The log-likelihood, neglecting constant terms, is proportional to

$$l = -\log |\mathbf{R} + \sigma \mathbf{d} \mathbf{d}^\dagger| - \mathrm{tr}\left[(\mathbf{R} + \sigma \mathbf{d} \mathbf{d}^\dagger)^{-1} \mathbf{S} \right] \quad . \tag{169}$$

Expanding the determinant as

$$\|\mathbf{R} + \sigma\mathbf{d}\mathbf{d}^\dagger\| = \|\mathbf{R}\|(1 + \sigma\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{d}) \tag{170}$$

and expanding the inverse of the covariance matrix using the Woodbury identity

$$(\mathbf{R} + \sigma\mathbf{d}\mathbf{d}^\dagger)^{-1} = \mathbf{R}^{-1} - \sigma\frac{\mathbf{R}^{-1}\mathbf{d}\mathbf{d}^\dagger\mathbf{R}^{-1}}{1 + \sigma\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{d}} \quad, \tag{171}$$

the terms of the log-likelihood that are functions of σ can be written

$$l = -\log(1 + \sigma\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{d}) + \sigma\frac{\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{d}}{1 + \sigma\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{d}} \quad. \tag{172}$$

Taking the derivative of this equation with respect to σ and setting the result to zero yields

$$\frac{dl}{d\sigma} = \frac{-\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{d}}{1 + \sigma\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{d}} + \frac{\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{d}}{1 + \sigma\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{d}} - \sigma\frac{\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{d}\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{d}}{(1 + \sigma\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{d})^2} = 0 \quad, \tag{173}$$

and the maximizing spectral weight is

$$\dot\sigma = \frac{1}{(\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{d})^2}\left[\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{d} - \mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{d}\right] \quad. \tag{174}$$

This will also satisfy the necessary condition on the second derivative of the log-likelihood. As this is assumed to be a spectral point, the contribution of the energy must be restricted to a positive quantity; therefore, the maximum-likelihood estimate of σ which satisfies this constraint is

$$\dot\sigma = \max(0, \dot\sigma) \quad. \tag{175}$$

When the estimated spectral weight is non-zero, then it can be shown that the increase in likelihood for the covariance estimate that contains this spectral estimate compared to the likelihood without this estimate is

$$\Delta l = l(\mathbf{Z}; \mathbf{B}, \mathbf{R} + \sigma\mathbf{d}\mathbf{d}^\dagger) - l(\mathbf{Z}; \mathbf{B}, \mathbf{R}) = \frac{\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{d}}{\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{d}} - 1 - \log\frac{\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{d}}{\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{d}} \quad. \tag{176}$$

Δl is a monotonic function of $(\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{d})/(\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{d})$ since this quantity is assumed to be greater than or equal to one $(\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{d} - \mathbf{d}^\cdot\mathbf{R}^{-1}\mathbf{d} \geq 0)$.

This estimator can be modified slightly by including an initial estimate of the spectral weight in $\mathbf{R}$. If the initial estimate is ν, then Equation (174) can be written as a function of $\mathbf{R}$ and the initial estimate so that

$$\dot{\sigma}(\mathbf{R}) = \dot{\sigma}(\mathbf{R} + \nu \mathbf{dd}^\dagger) + \nu \qquad . \tag{177}$$

Making the restriction that $\sigma + \nu$ be positive, the maximum-likelihood estimate of σ using the initial estimate ν is

$$\hat{\sigma} = \max(0, \dot{\sigma} + \nu) \qquad . \tag{178}$$

This property will be utilized in the proposed covariance estimator.

It is interesting that the maximum-likelihood estimator for a single point contains those terms that are used in the update equation for the spectral estimates in the EM iterations of Chapter 6. The difference between the update equation and this estimator is that the term added to the current estimate in the EM iteration is weighted by the square of the current spectral estimate; while here, the estimate is weighted by the square of the Capon spectral estimator.

7.2 Description of Variable Basis Estimator

In this section an iterative algorithm is defined that can be used to estimate the covariance matrices satisfying the spatial constraint. The likelihood produced by the resulting sequence of estimates is a non-decreasing function, and stable points of the estimator satisfy the conditions necessary for a maximum-likelihood estimator.

This algorithm is based on the estimator of Chapter 6 but does not require the large number of terms that is used to represent the entire space of constrained covariance matrices in order to represent the estimate. The data model assumed here is the same as that used in the previous chapters. P mutually independent data samples are available; the first G are assumed to be non-zero mean primary data $\mathbf{z}_1 \ldots \mathbf{z}_G$, with density $N(b_k \mathbf{d}, \mathbf{R})$, and the last K are the secondary data $\mathbf{z}_{G+1} \ldots \mathbf{z}_P$, with density $N(\mathbf{0}, \mathbf{R})$.

The estimation algorithm consists of the five steps that follow.

1. Form an initial estimate of the sample covariance matrix as

$$\mathbf{S} = \frac{1}{P} \left[\sum_{k=1}^{G} (\mathbf{z}_k - \hat{b}_k \mathbf{d})(\mathbf{z}_k - \hat{b}_k \mathbf{d})^\dagger + \sum_{k=G+1}^{P} \mathbf{z}_k \mathbf{z}_k^\dagger \right] \qquad . \tag{179}$$

where $\hat{b}_k = \mathbf{d}^\dagger \mathbf{z}_k / \mathbf{d}^\dagger \mathbf{d}$. These are the estimates of b that result from the assumption that the noise covariance matrix is a diagonal matrix. The term of the sample covariance matrix corresponding to the primary data can also be written as

$$\sum_{k=1}^{G} (\mathbf{z}_k - b_k \mathbf{d})(\mathbf{z}_k - b_k \mathbf{d})^\dagger = \left(\mathbf{I} - \frac{\mathbf{d}\mathbf{d}^\dagger}{\mathbf{d}^\dagger \mathbf{d}} \right) \sum_{k=1}^{G} \mathbf{z}_k \mathbf{z}_k^\dagger \left(\mathbf{I} - \frac{\mathbf{d}\mathbf{d}^\dagger}{\mathbf{d}^\dagger \mathbf{d}} \right) \qquad . \qquad (180)$$

It will be shown in the following chapter that the terms in Equation (179) are sufficient statistics for the estimate of the covariance matrix.

2. Form an initial set of L basis functions and weights. It is assumed that no prior knowledge of the interference directions exists. Simulations have shown that the number of iterations required to achieve a given likelihood is dependent upon the initial estimate; therefore, initialization is discussed in detail here. Two methods are proposed for this initialization. For both of these methods, if the constraint cone has a hyperplane boundary as shown in Chapter 5, then the bases for this boundary should be in the initial set of bases. Also, if the variance of the receiver noise is unknown and must be estimated, the elementary vectors will be used to form N of these basis functions.

- The first method begins with L bases that are parameterized by a uniform distribution of angles throughout space. The direction vectors corresponding to these basis functions are the columns of the matrix $\mathbf{D}$, which should be full rank. The spectral weights on these direction vectors are set to a value greater than zero but smaller than the expected contribution of energy due to receiver noise or due to spatially distributed noise for the region near that direction. With these spectral weights and basis functions, form an estimate of the covariance matrix.

  Using the set of basis functions, search for the basis function whose weight, if increased, would result in the largest increase in likelihood. This is the basis function generated by the vector that maximizes Equation (176) or equivalently

  $$\mathbf{d}_k = \operatorname*{argmax}_{\mathbf{d} \in \mathbf{D}} \frac{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{S} \mathbf{R}^{-1} \mathbf{d}}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d}} \qquad . \qquad (181)$$

  For this basis function and current weight c_k, calculate the new spectral weight

  $$\dot{\sigma}_k = \max \left(\sigma_k, \sigma_k + \frac{1}{(\mathbf{d}_k^\dagger \mathbf{R}^{-1} \mathbf{d}_k)^2} \left[\mathbf{d}_k^\dagger \mathbf{R}^{-1} \mathbf{S} \mathbf{R}^{-1} \mathbf{d}_k - \mathbf{d}_k^\dagger \mathbf{R}^{-1} \mathbf{d}_k \right] \right) \qquad . (182)$$

  Using this spectral weight, update the inverse of the covariance matrix using the Woodbury formula, an $O(N^2)$ operation, and update the term $\mathbf{R}^{-1} \mathbf{S} \mathbf{R}^{-1}$. Using the Woodbury formula, this can also be implemented as an $O(N^2)$ operation. Repeat this estimation of initial spectral weights until L weights have been set or until the estimated weight for the basis function with the maximum increase in likelihood results in a decrease or no change in its initial weight. The same weight will not be updated consecutively. The estimate is linear with respect to the spectral weight, and the gradient has been set to zero by the estimated weight.

- The second initialization method begins similarly to the first, by initializing part of the directions of arrival of the spectral points uniformly throughout space with a small weight; however, it then departs by allowing the directions for the other bases to vary. Using $L/2$ fixed vectors and their initial weights, form an estimate of the covariance matrix $\mathbf{R}$. With this initial estimate of the covariance, the remaining directions of arrival and weights will be found by using the conditional maximum-likelihood estimate for the weights. Rather than maximizing the initial likelihood using the fixed bases, the directions of arrival for the other $L/2$ basis functions are varied so as to maximize the likelihood.

Beginning with the initial estimate of $\mathbf{R}$, perform a search for the direction of arrival that would result in the maximum increase in likelihood. This is the direction (θ_k, φ_k) that satisfies Equation (176) or

$$(\theta_k, \varphi_k) = \underset{\theta, \varphi}{\text{argmax}} \ \frac{\mathbf{d}^\dagger(\theta, \varphi)\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{d}(\theta, \varphi)}{\mathbf{d}^\dagger(\theta, \varphi)\mathbf{R}^{-1}\mathbf{d}(\theta, \varphi)} \quad . \tag{183}$$

This is a computationally intensive task requiring on the order of N^2 computations per point in the search. This can be performed by using an array of fixed-basis vectors similar to the array used in the estimator of Chapter 6. After the direction of arrival for the maximum increase in likelihood is found by this coarse search, the direction can be refined by the use of a search between the directions of arrival of the array of basis vectors until a suitable *stopping criterion, such as the* difference in the angle of arrival is within some ε, is reached. The weight for this direction can be found by the conditional maximum-likelihood estimate as

$$\dot{\sigma} = \max(0, \frac{1}{(\mathbf{d}_k^\dagger \mathbf{R}^{-1} \mathbf{d}_k)^2} \left[\mathbf{d}_k^\dagger \mathbf{R}^{-1} \mathbf{S} \mathbf{R}^{-1} \mathbf{d}_k - \mathbf{d}_k^\dagger \mathbf{R}^{-1} \mathbf{d}_k \right]) \quad . \tag{184}$$

After estimating a direction of arrival and weight, the inverse of the initial covariance estimate and the matrix product $\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}$ are updated, and the process is continued until all L directions and weights have been found. If during this process the weight for the direction vector corresponding to the maximum increase in likelihood is negative, the remaining direction vectors should be distributed as uniformly throughout space as possible and the weights set to a small value.

For both of these initialization methods, the initial set of bases is chosen so as to form a linearly independent set. The bases for all iterations are restricted to form a linearly independent set. The resulting set of bases forms an L-dimensional constraint cone. With this restriction, there will not be two bases parameterized by the same direction of arrival. This allows the estimate to move in any direction interior to the constraint region during the estimation procedure.

3. Using this set of basis functions and initial weights, perform EM iterations using the method of Chapter 6. These equations are repeated here:

$$\hat{\boldsymbol{\Sigma}}_{jj}^{p+1} = \left[\hat{\boldsymbol{\Sigma}}^p + \hat{\boldsymbol{\Sigma}}^p \mathbf{D}^\dagger \left(\hat{\mathbf{R}}^{p^{-1}} \mathbf{S}^p \hat{\mathbf{R}}^{p^{-1}} - \hat{\mathbf{R}}^{p^{-1}}\right) \mathbf{D}\hat{\boldsymbol{\Sigma}}^p\right]_{jj} \quad , \qquad (185)$$

$$\hat{\mathbf{R}}^{p+1} = \mathbf{D}\hat{\boldsymbol{\Sigma}}^{p+1}\mathbf{D}^\dagger \quad , \qquad (186)$$

and

$$\hat{b}_k^p = \frac{\mathbf{d}_1^\dagger \hat{\mathbf{R}}^{-1} \mathbf{z}_k}{\mathbf{d}_1^\dagger \hat{\mathbf{R}}^{-1} \mathbf{d}_1} \quad , \qquad (187)$$

where the sample covariance matrix is

$$\mathbf{S}^p = \frac{1}{P}\left[\sum_{k=1}^{G}(\mathbf{z}_k - \hat{b}_k^p \mathbf{d}_1)(\mathbf{z}_k - \hat{b}_k^p \mathbf{d}_1)^\dagger + \sum_{k=G-1}^{P} \mathbf{z}_k \mathbf{z}_k^\dagger\right] \quad . \qquad (188)$$

The direction vector corresponding to the mean does not need to be one of the basis vectors that is used to estimate the covariance matrix because the modification to the estimator of Chapter 6 maximizes the likelihood in each iteration separately with respect to the mean.

4. If the estimate is moving outside of the polyhedral cone formed by the initial set of basis functions, the estimate will approach the boundary of the constraint set. This is determined by examining the conditional maximum-likelihood spectral estimates given the current estimate of the covariance matrix. If the conditional maximum-likelihood estimate would be zero using the modified form for the conditional maximum-likelihood estimate found in Equation (178), then the estimate is approaching a hyperplane boundary. Examining the approach of the estimate to a hyperplane boundary uses only those quantities that need to be calculated to update the spectrum estimates using the EM algorithm and requires one additional multiply and a divide per point. Examining the approach to a hyperplane boundary should be accomplished before the update of the spectrum estimates. If any conditional spectrum estimate would be zero, then the basis function and weight should be removed and replaced with a basis function that when added would result in the maximum increase in likelihood using the procedure outlined in the initialization, that is, with the basis function corresponding to the angle of arrival that maximizes

$$(\theta, \varphi) = \underset{\theta, \varphi}{\operatorname{argmax}} \frac{\mathbf{d}^\dagger(\theta, \varphi)\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{d}(\theta, \varphi)}{\mathbf{d}^\dagger(\theta, \varphi)\mathbf{R}^{-1}\mathbf{d}(\theta, \varphi)} \quad . \qquad (189)$$

If there is not a different basis function that would result in an increase in likelihood, then the current basis function and weight are retained. It has been found that replacing a single basis function and weight in any iteration is sufficient. Additionally, it has been found that the examination of the approach to a hyperplane boundary

does not need to be performed for each iteration since the estimate will be interior to the new polyhedral cone and can move within this cone.

5. Perform steps 3 and 4 until a suitable stopping criterion is reached. The stopping criterion could be based on the change in likelihood, or it could be based on performing a fixed number of iterations.

7.3   Estimator Properties

In Chapter 6 it was shown that under certain conditions a maximum-likelihood estimator for the mean and the covariance may not exist for this model. There will typically be a maximal proper subspace orthogonal to the direction of arrival for the mean, as discussed in Section 6.3. If all of the data are assumed to have a non-zero mean and the sample covariance matrix is not full rank, the likelihood will be unbounded above and the maximum-likelihood estimate will not exist.

For a non-singular sample covariance matrix and data that are assumed to be zero mean, the maximum-likelihood estimate will exist [28]. If the sample covariance matrix is singular, then the existence of the maximum-likelihood estimate will be determined by the array geometry, and the dimension of the constraint space and general statements will not be made here.

To ensure that a maximum-likelihood estimate exists, receiver noise with a non-zero variance should be included in the model. This adds to the covariance matrix during each iteration just as it did in Chapter 6.

A stable point of this algorithm satisfies the necessary conditions for a maximizer of the likelihood. If the necessary conditions on the gradient are satisfied, then the sequence of estimates produced by the method of Chapter 6 will be stable as the gradient of the likelihood appears directly in the update equation. This was pointed out in Chapter 6. If these estimates are stable, then the estimate is interior to the finite bases constraint cone, and no additional bases will be added. If the gradient is positive in any direction interior to the constraint set, then the sequence of estimates will move in that direction until the estimate is stable or until the estimate approaches a hyperplane and an additional basis dyad is added. Since a new basis element, when added, maximizes the likelihood, the estimate will only move toward a maximizer of the likelihood.

7.4   Comparisons to Fixed-Bases Estimator

7.4.1   Convergence

The estimation procedures presented in this chapter appear to be much more complex than the fixed-bases estimation procedure. In this section we show that, although the variable basis estimator is theoretically more complex, the computational requirements can be fewer than the requirements for the fixed-bases estimator.

The typical environment for adaptive beamforming and detection can have a few spatially discrete interference sources with energy levels that are 50–80 dB higher than the receiver noise

[51]. With levels this high, the true covariance matrix will lie near the boundary of the constraint region. In order to adequately represent these matrices, the fixed-bases estimator must use a large number of terms in the finite sum and accept a small loss in detection and beamformer performance or use a small number of terms and accept a larger loss in performance. This performance trade-off is not required if the basis functions are allowed to vary.

Reducing the number of terms required to represent the estimate can result in a more rapid approach of the estimate to the maximum-likelihood solution. Recalling the update equations for the spectral estimates of Equation (167)

$$\sigma_j^{p+1} = \sigma_j^p (1 + \sigma_j^p \Delta) \tag{190}$$

for a particular covariance matrix, if the energy is spread among a larger number of spectral weights, then the estimate will not move as far in the direction of the gradient. Assume that the covariance estimate has an error term $\mathbf{R}_\epsilon^p$ that is approaching zero. If this term is formed by a single spectral weight, then

$$\mathbf{R}_\epsilon^p = \sigma^p \mathbf{d}\mathbf{d}^\dagger \qquad . \tag{191}$$

If this term is represented by a large number of spectral weights, then we have

$$\mathbf{R}_\epsilon^p = \sum_{g=1}^{G} \sigma_g^p \mathbf{d}_g \mathbf{d}_g^\dagger \qquad . \tag{192}$$

It will be assumed that the array geometry is such that $\sum_{g=1}^{G} \sigma_g^p = \sigma^p$, and the spectral weights are evenly distributed. $\mathbf{R}_\epsilon^p$ is nearly the same in either case. This could occur if the direction vectors forming the basis functions are closely spaced. Then $\sigma_g^p = \sigma^p / G$. It will also be assumed that the gradient of the likelihood is equal for all of the terms. Then

$$\mathbf{R}_\epsilon^{p+1} = \sum_{g=1}^{G} \frac{\sigma^p}{G} (1 + \Delta \frac{\sigma^p}{G}) \mathbf{d}_g \mathbf{d}_g^\dagger \tag{193}$$

for the estimate containing many terms, and

$$\mathbf{R}_\epsilon^{p+1} = \sigma^p (1 + \Delta \sigma^p) \mathbf{d}_g \mathbf{d}_g^\dagger \tag{194}$$

for the estimate with a single term. The fractional change in the first case is $\Delta \sigma^p / G$, and for the second it is $\Delta \sigma^p$, showing that with the energy spread among many directions will result in slowing

the rate of convergence. $\mathbf{R}_e$ is approaching zero, so the estimate cannot go too far in the direction of the gradient in either case.

7.4.2 Computational Complexity

The algorithm proposed in this chapter has lower computational requirements than the fixed-bases estimator. The computational burden will be investigated in terms of the number of multiplications required. The number of terms in the fixed-bases finite sum is a quantity that can be traded off against the performance of detection and beamforming algorithms using the estimates produced. The number of bases is a multiple of the dimension of the constraint space, for example, Q. The differences in the number of multiplications will occur for the spectral weight updates and generating the covariance from the spectral weights. The variable basis method will also have the overhead of determining the weights when adding or reducing the number of terms in the spectrum. The quadratic term $\mathbf{d}^{\dagger}(\cdots)\mathbf{d}$ of the spectrum update requires $N^2 + O(N)$ multiplications for each spectral point in each iteration. For the fixed-bases estimator, this will be $QLN^2 + O(N)$. The number of multiplications for the method discussed in this chapter for the spectral update is $LN^2 + O(N)$.

To calculate a new covariance matrix from the spectrum samples requires $N^2/2 + O(N)$ multiplies for each spectral point. For the fixed-bases estimator the number per iteration will then be $QLN^2/2 + O(N)$, and for the variable basis, $LN^2/2 + O(N)$.

Each time a basis function is added or subtracted, a search for the maximum increase in likelihood is performed. The most intensive task is finding the approximate location of the maximum, which, if performed using fixed bases, would require approximately $QLN^2 + O(N)$ multiplies. After this, increasing the resolution using a binary search may require an additional $10N^2$ multiplies. Updating the inverse of the covariance matrix and gradient of the likelihood requires $3N^2$ operations plus $3N^2$ for the removal of the weight that is approaching zero. Additionally, the updates for the spectral estimates must be recalculated requiring $LN^2 + O(N)$ multiplies. It is assumed that a basis function must be added every J iterations; therefore, the number of multiplications per iteration is $(QL + L + 16)N^2/J + O(N)$. For each method, the number of real multiplications per iteration, ignoring lower order terms, is

$$
\begin{array}{ll}
\text{fixed basis} & 6QLN^2 \\
\text{variable basis} & 6LN^2 + 4(QL + L + 16)N^2/J
\end{array}
\qquad .
$$

Looking at the relative values of these quantities, the terms Q, N and J are all about the same magnitude. Assuming that $QL >> L + 16$, the fixed-bases estimator will require $O(LN^3)$ multiplies, while the variable basis method will require $O(LN^2)$. The variable basis method actually requires less computation than the fixed-bases method.

7.5 Simulation Results

A few brief simulations will be used to compare the estimator of this chapter to the fixed-bases estimator of Chapter 5. A 4-element minimum redundancy linear array has been simulated. The dimension of the space of matrices is 13. The fixed-bases estimator uses 26, 130, and 900 direction vectors to form the basis. All of the estimators also estimated a diagonal term that was constrained to be identically distributed for each of the vector elements. The variable basis estimator of this chapter used a fixed resolution of 0.05 degree or the equivalent of 3600 basis vectors. The true covariance was used in place of the sample covariance so that the maximum log-likelihood was known, and this was subtracted from the likelihoods before plotting so that the maximum for each of the plots is 0. The need to adjust the bases was determined every fourth iteration for the method of this chapter. Two interference environments are simulated. The first environment consists of receiver noise with a variance of 1, uniform spatial interference with a variance of 1 for a single vector element (which contributes an identity term here due to simulated sensor placement), and three discrete interference sources located at azimuths of -25.78, 16.04, and 34.95 degrees with power levels of 30, 40, and 60 dB across the array, respectively, when compared to the receiver noise. The second environment consists only of the receiver noise and uniform spatial interference, and the covariance matrix is interior to the polyhedral cone formed by the set of bases for each of the fixed-bases estimators.

A plot of the likelihood versus iteration is shown here in Figure 17. We see that the variable bases estimator has a more rapid rate of convergence and results in a final likelihood that is greater than the fixed-bases estimator. Iterations where basis functions were added are shown by the increase in the rate of convergence. The estimator of this chapter re-formed the set of bases 15 times during these iterations for the first interference set and did not require changing the bases for the second. The estimators of Chapter 6 were initialized by a constant spectrum value (all ones). The difference in the convergence rate is quite dramatic when comparing the method of this chapter to the estimator of the previous chapter.

To show that the performance difference is not due solely to the new initialization method, the estimator of Chapter 6 was initialized by the first method described in this chapter for Figure 18. Although this initialization results in a higher initial likelihood, the estimator with the large number of terms (with some spectral weights approaching zero or with the small number of terms that cannot represent the true covariance matrix) does not converge after 1000 iterations. One additional plot, Figure 19, will be used simply to expand the scale of the second plot in Figure 18. This shows that the initialization causes the estimate to be very nearly the maximum-likelihood estimate for this noise environment and the small number of bases used.

For the second interference environment, all of the fixed-bases estimators are able to represent the covariance matrix; however, for the estimator with 900 bases the convergence is so slow that the estimator has not converged after 1000 iterations.

These plots show the trade-off between the number of bases of the previous chapter and the ability to represent strong discrete interference sources with the rate of convergence.
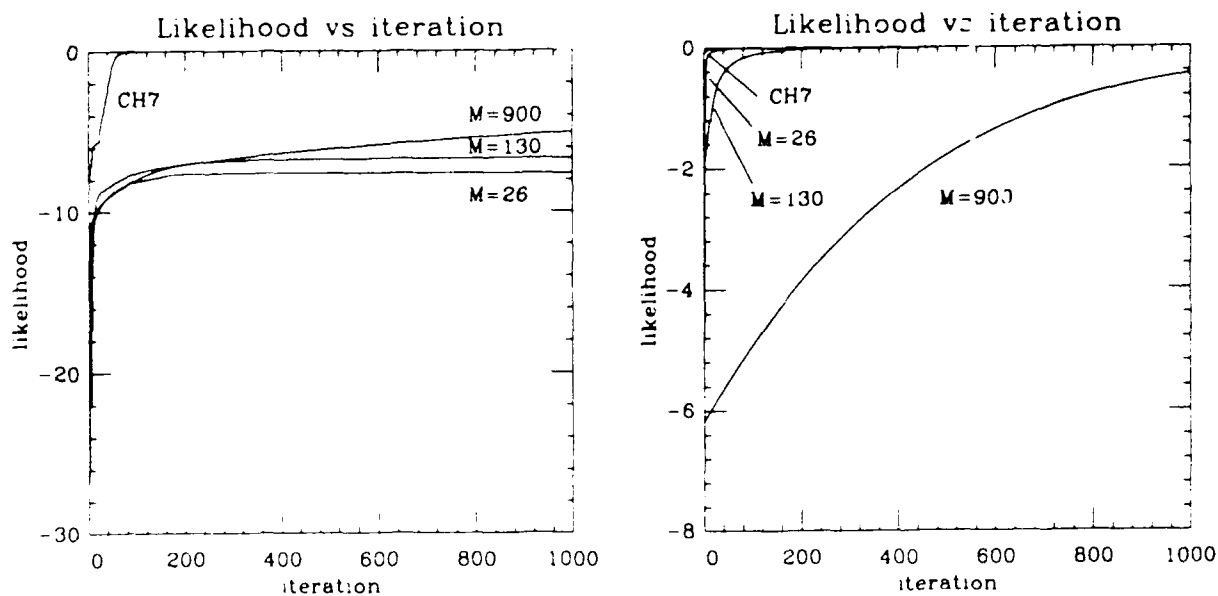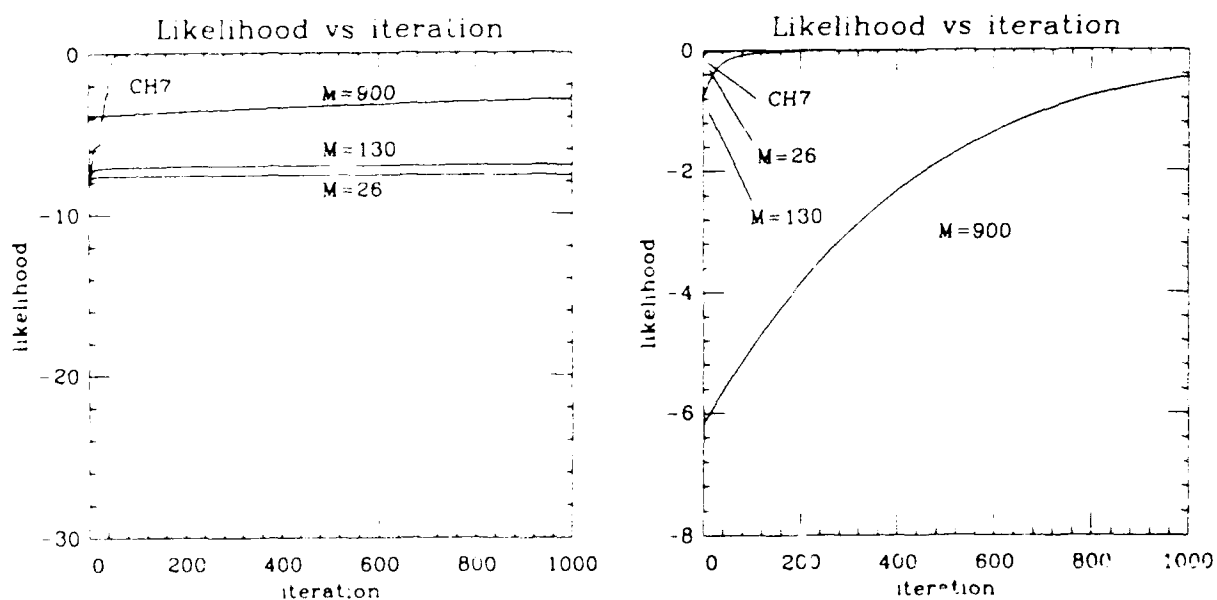
Figure 17.   Comparison of convergence rates.

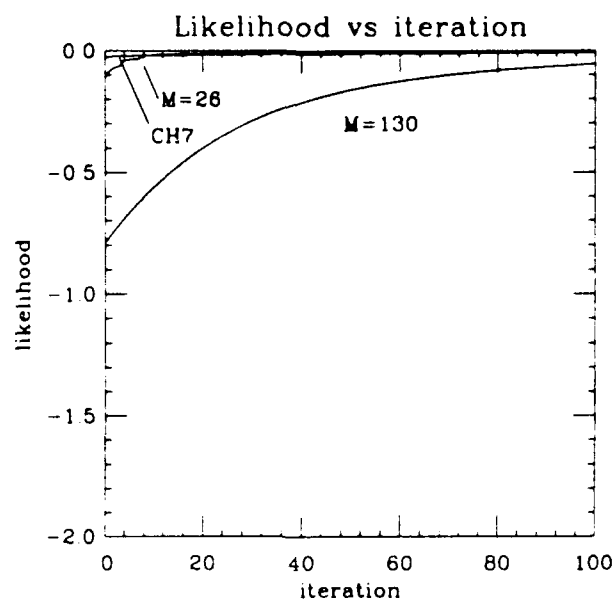Figure 18.   Comparison of convergence rates using a better initialization.

88

Figure 19.   Comparison of convergence rates expanded scale.

7.6   Conclusion

In Chapter 5 the maximal representation of covariance matrices was discussed. In this chapter, an estimation method has been presented that is based on the maximal representation. This method has been shown to have computational and convergence benefits compared to the fixed-basis finite representation estimator developed in Chapter 6.

In the following chapter, we will use the covariance estimates and the likelihood found through this procedure to develop adaptive detection and beamforming methods.

8. BEAMFORMING AND DETECTION METHODS

The material presented in the three previous chapters focused on estimating the parameters that are indicated by the signal model without explicitly stating how the results of these estimation algorithms would be used. In this chapter, several beamforming and detection methods that make use of these parameter estimates and estimation algorithms are presented.

8.1 Beamforming

In this section, methods of estimating the mean corresponding to a given direction are proposed. These beamforming methods are based on the approaches discussed in Chapter 3 and use maximum-likelihood techniques. It is assumed that a single primary vector z may have a non-zero mean of the form $b\mathbf{d}$, and K mutually independent secondary vectors are available that are subject to a density function with the same covariance as the primary vector. The object of the beamformer is to estimate the complex amplitude b.

1. The first method is to perform joint estimation of the mean and the covariance matrix by the maximum-likelihood estimation procedures introduced in Chapters 6 and 7. These procedures will use all of the available data to form the estimates of the mean and the covariance.

2. The second method is similar to that suggested by Reed et al. [10]: Substitute the constrained estimate of the covariance using only the secondary data for the covariance matrix in the maximum signal-to-noise ratio beamformer

$$\mathbf{w} = k\mathbf{R}^{-1}\mathbf{d} \qquad . \qquad (195)$$

Additionally, the constant k is set so that the gain in the desired look direction is equal to unity. The weight vector is then

$$\mathbf{w} = \frac{\hat{\mathbf{R}}^{-1}\mathbf{d}}{\mathbf{d}^{\dagger}\hat{\mathbf{R}}^{-1}\mathbf{d}} \qquad . \qquad (196)$$

This weight vector will be used to form an estimate of the mean in the primary vector z by

$$\hat{b} = \mathbf{w}^{\dagger}\mathbf{z} \qquad . \qquad (197)$$

For unconstrained maximum-likelihood covariance estimates, these two procedures produce equivalent estimates of the mean. This property has not been proven to be true for constrained covariance estimates.

Beamformers based on constrained covariance estimates will be compared to the beamformer using unconstrained maximum-likelihood covariance estimates and to the optimum linear beamformer formed when the covariance matrix is known. Because closed-form expressions for the covariance estimates and hence the mean estimates have not been found, the beamforming methods will be compared through simulations.

Performance will be evaluated by comparing the bias and variance of the estimates produced by each method. This also compares the variance of the estimates to the Cramer-Rao bound when the covariance is known since the known covariance maximum-likelihood estimator is efficient.

The covariance estimates will be evaluated in terms of their application to beamforming by comparing the signal-to-noise ratio loss factor introduced by Reed et al. [10] and discussed in Chapter 3. The loss factor is

$$\rho = \frac{|\mathbf{w}^\dagger \mathbf{d}|^2}{\mathbf{d}^\dagger \mathbf{R}^{-1}\mathbf{d}\,\mathbf{w}^\dagger \mathbf{R}\mathbf{w}}$$

$$= \frac{|\mathbf{d}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{d}|^2}{\mathbf{d}^\dagger \mathbf{R}^{-1}\mathbf{d}\,\mathbf{d}^\dagger \hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{d}} \quad . \tag{198}$$

This loss factor will verify the comparison of the variance since, for an unbiased estimator, minimizing the variance is equivalent to maximizing the signal-to-noise ratio. The loss factor is a random variable and will be presented as a histogram to indicate the corresponding density function.

Additionally, the spatial response of the beamforming methods will be compared. This is presented simply to show the variance in the spatial response produced by each of the beamforming methods.

8.2 Adaptive Detection

The proposed detectors assume that there is a single primary vector $\mathbf{z}$, which may have a non-zero mean (G=1) of the form $b\mathbf{d}$. $\mathbf{d}$ is known, and b is assumed to be unknown. K additional mutually independent secondary vectors $\mathbf{Z} \equiv \{\mathbf{z}_1 \cdots \mathbf{z}_K\}$ are also available, and these vectors share the same covariance matrix $\mathbf{R}$.

Two methods of utilizing the algorithms presented in Chapters 6 and 7 are introduced.

1. The first detection method is to utilize the generalized likelihood ratio test procedure [2] given by

$$\Lambda(\mathbf{z}, \mathbf{Z}) = \frac{\max\limits_{b,\mathbf{R}|H_1} f(\mathbf{z}, \mathbf{Z}; b, \mathbf{R}|H_1)}{\max\limits_{\mathbf{R}|H_0} f(\mathbf{z}, \mathbf{Z}; \mathbf{R}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \quad . \tag{199}$$

The test statistic for this method cannot be written in closed form; rather, the likelihood ratio must be computed explicitly from the maximum-likelihood estimation procedures under each hypothesis.

2. The constrained maximum-likelihood estimate of the covariance matrix based on the secondary data may be substituted for the known covariance in the unknown mean, known covariance generalized likelihood ratio test. This can be justified if the

estimate of the covariance matrix is estimated with sufficient accuracy. The resulting test is

$$\frac{|\mathbf{d}^\dagger \hat{\mathbf{R}}^{-1} \mathbf{z}|^2}{\mathbf{d}^\dagger \hat{\mathbf{R}}^{-1} \mathbf{d}} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \quad . \tag{200}$$

For both of these methods, the test statistics are positive. In the simulations discussed in the following chapter, a monotonic function of the test statistics is plotted.

Without a closed-form expression for the test statistics under these methods, deriving density functions and evaluating performance based on these density functions cannot be accomplished. However, some of the properties of these detectors will be discussed in the next section. How these detectors perform compared to the other detectors will be evaluated by simulations in the next chapter.

8.3 Discussion

8.3.1 Beamforming

Because closed-form expressions for the constrained covariance estimates have not been found, an explicit formula for the loss in signal-to-noise ratio has not been derived. Other properties of the beamformers can be proven. For the first beamformer, the sufficient statistic for the estimate of the covariance matrix only involves the secondary data and the primary data that are orthogonal to the subspace spanned by the known direction vector. In addition, in the vector space of matrices, the primary data will only affect the estimated covariance matrix in the subspace orthogonal to $\mathbf{dd}^\dagger$. It can be shown that the second beamformer provides an unbiased estimate of the mean.

The independence of the estimated covariance matrix to that portion of the primary data in the span of the direction vector can be shown by writing the log-likelihood and then maximizing with respect to the amplitude of the mean b. Substituting the maximum-likelihood estimate of b, the log-likelihood without simplifying is

$$l = -NP \log \pi - P \log \|\mathbf{R}\| - \sum_{k=1}^{G} (\mathbf{z}_k - \frac{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{z}_k}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d}} \mathbf{d})^\dagger \mathbf{R}^{-1} (\mathbf{z}_k - \frac{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{z}_k}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d}} \mathbf{d}) - \sum_{k=G+1}^{P} \mathbf{z}_k^\dagger \mathbf{R}^{-1} \mathbf{z}_k \quad . \tag{201}$$

The terms corresponding to data vectors that are non-zero mean can be written

93

$$\left(\mathbf{z}_k - \frac{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{z}_k}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d}} \mathbf{d} \right)^\dagger \mathbf{R}^{-1} \left(\mathbf{z}_k - \frac{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{z}_k}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d}} \mathbf{d} \right)$$

$$= \mathbf{z}_k^\dagger \mathbf{R}^{-\frac{1}{2}} \left(\mathbf{I} - \frac{\mathbf{R}^{-\frac{1}{2}} \mathbf{dd}^\dagger \mathbf{R}^{-\frac{1}{2}}}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d}} \right) \left(\mathbf{I} - \frac{\mathbf{R}^{-\frac{1}{2}} \mathbf{dd}^\dagger \mathbf{R}^{-\frac{1}{2}}}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d}} \right) \mathbf{R}^{-\frac{1}{2}} \mathbf{z}_k$$

$$= \mathbf{z}_k^\dagger \mathbf{R}^{-\frac{1}{2}} \left(\mathbf{I} - \frac{\mathbf{R}^{-\frac{1}{2}} \mathbf{dd}^\dagger \mathbf{R}^{-\frac{1}{2}}}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d}} \right) \mathbf{R}^{-\frac{1}{2}} \mathbf{z}_k \qquad . \tag{202}$$

The terms in the braces $(\cdot)$ on the right are projection matrices with a null space for components in the primary vector in the subspace spanned by $\mathbf{d}$. This can be clearly seen by the equality

$$\mathbf{z}_k^\dagger \mathbf{R}^{-\frac{1}{2}} \left(\mathbf{I} - \frac{\mathbf{R}^{-\frac{1}{2}} \mathbf{dd}^\dagger \mathbf{R}^{-\frac{1}{2}}}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d}} \right) \mathbf{R}^{-\frac{1}{2}} \mathbf{z}_k$$

$$= \mathbf{z}_k^\dagger \left(\mathbf{I} - \frac{\mathbf{dd}^\dagger}{\mathbf{d}^\dagger \mathbf{d}} \right) \mathbf{R}^{-\frac{1}{2}} \left(\mathbf{I} - \frac{\mathbf{R}^{-\frac{1}{2}} \mathbf{dd}^\dagger \mathbf{R}^{-\frac{1}{2}}}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d}} \right) \mathbf{R}^{-\frac{1}{2}} \left(\mathbf{I} - \frac{\mathbf{dd}^\dagger}{\mathbf{d}^\dagger \mathbf{d}} \right) \mathbf{z}_k \qquad . \tag{203}$$

The result is that the maximum-likelihood estimate of the covariance matrix only depends on the vector components of the non-zero mean data in the subspace orthogonal to $\mathbf{d}$. From this it can be seen that the sufficient statistic for the estimate of the covariance matrix is the sum of the outer products of the zero mean data and the sum of the outer products of the vector components of the non-zero mean data in the subspace orthogonal to $\mathbf{d}$.

A property of the estimates is that, over the vector space of matrices, the primary data will not affect the covariance estimates for variations of the form $\delta \mathbf{R} = \sigma \mathbf{dd}^\dagger$. The likelihood for variations of this type is not affected by the primary data; the primary data will only affect the estimate of the covariance matrix for directions that are orthogonal to $\mathbf{dd}^\dagger$. The necessary condition that must be satisfied by maximum-likelihood covariance estimates interior to the constraint region is

$$\delta l = \text{tr}\left[(\mathbf{R}^{-1} \mathbf{S} \mathbf{R}^{-1} - P \mathbf{R}^{-1}) \delta \mathbf{R} \right] -$$

$$\sum_{k=1}^{G} (\mathbf{z}_k - \frac{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{z}_k}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d}} \mathbf{d})^\dagger \mathbf{R}^{-1} \, \delta \mathbf{R} \, \mathbf{R}^{-1} (\mathbf{z}_k - \frac{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{z}_k}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d}} \mathbf{d}) = 0 \tag{204}$$

where $\mathbf{S} = \mathbf{ZZ}^\dagger$. Now let $\delta \mathbf{R} = \delta \sigma \mathbf{dd}^\dagger$. The resulting equation,

$$\delta l = \mathbf{d}^\dagger (\mathbf{R}^{-1} \mathbf{S} \mathbf{R}^{-1} - P \mathbf{R}^{-1}) \mathbf{d} \delta \sigma \qquad , \tag{205}$$

94

does not contain the primary data. Because the primary data does not affect the likelihood for variations of this form, then this data will not affect the estimate of the covariance matrix in the subspace spanned by $\mathbf{dd}^\dagger$. This can also be shown by letting $\mathbf{R} = \mathbf{R}_o + \sigma\mathbf{dd}^\dagger$, assuming that $\mathbf{R}_o$ is non-singular. Substituting this expression in Equation (201), using the Woodbury formula and simplifying, the log-likelihood terms corresponding to the non-zero mean data are of the form

$$-\mathbf{z}_k^\dagger \mathbf{R}_o^{-1} \mathbf{z}_k + \frac{|\mathbf{d}^\dagger \mathbf{R}_o^{-1} \mathbf{z}_k|^2}{\mathbf{d}^\dagger \mathbf{R}_o^{-1} \mathbf{d}} \quad . \tag{206}$$

σ drops out of the log-likelihood expression for the non-zero mean terms, and the estimate of σ is indeterminate with respect to the primary data. This was discussed in a different context in Chapter 6 where it was shown that a maximum-likelihood estimator of the covariance matrix may not exist when all of the data are assumed to contain a mean.

The error in the beamformer plots used to illustrate the variance of the estimators can be related to the covariance matrix estimation error $\mathbf{E}$. If the true and estimated covariance matrices are $\mathbf{R}$ and $\hat{\mathbf{R}}$, then

$$\hat{\mathbf{R}} = \mathbf{R} + \mathbf{E} \quad . \tag{207}$$

Since $\mathbf{E}$ is Hermitian, then it is unitarily similar to a diagonal matrix and has a decomposition

$$\mathbf{E} = \mathbf{U}\Sigma\mathbf{U}^\dagger \quad , \tag{208}$$

where Σ is diagonal but not necessarily positive. The beamformer that results from using this estimated covariance matrix, neglecting the constant scale factor, is

$$\mathbf{w} = \hat{\mathbf{R}}^{-1}\mathbf{d} = (\mathbf{R} + \mathbf{U}\Sigma\mathbf{U}^\dagger)^{-1}\mathbf{d} \quad . \tag{209}$$

Using the Woodbury inversion formula [36], the weight vector is

$$\begin{aligned}
\mathbf{w} &= \mathbf{R}^{-1}\mathbf{d} - \mathbf{R}^{-1}\mathbf{U}(\mathbf{U}^\dagger\mathbf{U} + \Sigma^{-1})^{-1}\mathbf{U}^\dagger\mathbf{R}^{-1}\mathbf{d} \\
&= \mathbf{w}_o + \mathbf{w}_e \quad , \tag{210}
\end{aligned}$$

and the magnitude squared of the array response is

$$|\mathbf{w}^\dagger\mathbf{d}(\theta,\phi)|^2 = |(\mathbf{w}_o + \mathbf{w}_e)^\dagger\mathbf{d}(\theta,\phi)|^2 = |\mathbf{w}_o^\dagger\mathbf{d}(\theta,\phi)|^2 + \mathbf{w}_e^\dagger\mathbf{d}(\theta,\phi)\mathbf{d}(\theta,\phi)^\dagger(\mathbf{w}_o + \mathbf{w}_e) \quad . \tag{211}$$

This is the optimal response perturbed by the error in the estimation of the covariance matrix that appears in the weight error vector $\mathbf{w}_e$. If this weight error vector is within the span of $\mathbf{w}_o$, then the error will not change the array sidelobe pattern but only the scaling, which would be incorporated into the normalization of the weight vector.

The proof that the second estimator is unbiased is based on the statistical independence of the primary and the secondary data vectors. The density function of the estimated covariance matrix is unknown; however, the estimation of the covariance matrix does not involve the primary data vector. Let the mapping of the secondary data to the estimated covariance be written as $\mathbf{\Theta}(\mathbf{Z})$, then the expected value of the estimate is

$$
\mathrm{E}\left\{\frac{\mathbf{d}^\dagger\mathbf{\Theta}(\mathbf{Z})\mathbf{z}}{\mathbf{d}^\dagger\mathbf{\Theta}(\mathbf{Z})\mathbf{d}}\right\} = b\mathrm{E}\left\{\frac{\mathbf{d}^\dagger\mathbf{\Theta}(\mathbf{Z})\mathbf{d}}{\mathbf{d}^\dagger\mathbf{\Theta}(\mathbf{Z})\mathbf{d}}\right\} = b \qquad , \tag{212}
$$

and the estimator is unbiased.

8.3.2 Detection

Although closed-form expressions cannot be found for the estimates of the covariance matrices, some properties of the detectors based on the constrained covariance estimates can be analyzed. It will be shown that both of the detectors are invariant to scaling or a change of coordinates of the data and are invariant to scaling of the direction vectors. Invariance of the test statistic to scaling of the data is necessary for a detector to be CFAR and is itself a weak CFAR property.

In either of the tests described earlier, the test statistic will be invariant to scaling of the signal direction vector. This is a desirable property since no explicit normalization has been assumed. Without this property, detection performance would depend on the scaling. This property can be seen by examining the test statistic for the constrained AMF test

$$
\frac{|\mathbf{d}^\dagger\hat{\mathbf{R}}^{-1}\mathbf{z}|^2}{\mathbf{d}^\dagger\hat{\mathbf{R}}^{-1}\mathbf{d}} \underset{H_0}{\overset{H_1}{\gtrless}} \alpha \qquad . \tag{213}
$$

Any scaling of the direction vectors will be canceled in the resulting test statistic.

For the generalized likelihood ratio test, this property is based on the form of the log-likelihood under H_1. After maximizing the log-likelihood with respect to the mean, the terms in the likelihood that are dependent upon the signal direction vector are of the form

$$
\frac{\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{z}}{\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{d}}\mathbf{d} \qquad . \tag{214}
$$

It can be seen that the direction vectors enter into this equation so that these terms are invariant to scaling, and the result is that the maximum of the likelihood will be independent of the scaling of the direction vector.

It will now be shown that the test statistics are invariant with respect to an invertible change of coordinates. An invertible change in coordinates will preserve the dimension of the constraint space but may not preserve the obvious structure of the covariance matrix, e.g., Toeplitz. Beginning with the log-likelihood of the numerator,

$$l \quad = \quad -NP \log \pi - P \log \|\mathbf{R}\| \tag{215}$$

$$- \sum_{k=1}^{G} (\mathbf{z}_k - \frac{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{z}_k}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d}} \mathbf{d})^\dagger \mathbf{R}^{-1} (\mathbf{z}_k - \frac{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{z}_k}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d}} \mathbf{d}) - \sum_{k=G+1}^{P} \mathbf{z}_k^\dagger \mathbf{R}^{-1} \mathbf{z}_k \quad ,$$

a necessary condition for the covariance matrix, interior to the constraint region, that maximizes the likelihood is

$$\delta l \quad = \quad \mathrm{tr} \left[(\mathbf{R}^{-1} \mathbf{S} \mathbf{R}^{-1} - P \mathbf{R}^{-1}) \delta \mathbf{R} \right] -$$

$$\sum_{k=1}^{G} (\mathbf{z}_k - \frac{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{z}_k}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d}} \mathbf{d})^\dagger \mathbf{R}^{-1} \delta \mathbf{R} \, \mathbf{R}^{-1} (\mathbf{z}_k - \frac{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{z}_k}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d}} \mathbf{d}) = 0 \quad , \tag{216}$$

where $\mathbf{S} = \sum_{k=G+1}^{P} \mathbf{z}_k \mathbf{z}_k^\dagger$. Based on this expression, invariance to the scale and to a change of coordinates can be shown. Assume for data set $\mathbf{Z}$ that the solution to the above equation is $\hat{\mathbf{R}}$. Now subject the data to an invertible transformation of coordinates so that $\mathbf{y}_k = \mathbf{V} \mathbf{z}_k$, $\mathbf{p} = \mathbf{V} \mathbf{d}$. For this new data set the covariance matrix will be a member of the constraint space

$$\Omega_{\mathbf{y}} = \{ \mathbf{R_y} : \mathbf{R_y} = \mathbf{V} \mathbf{R_z} \mathbf{V}^\dagger, \ \forall \mathbf{R_z} \in \Omega_{\mathbf{x}} \} \quad . \tag{217}$$

It can be shown that a solution to Equation (216) will be $\mathbf{V} \hat{\mathbf{R}} \mathbf{V}^\dagger$ for data set $\mathbf{Y}$. Substituting this into Equation (216), all of the $\mathbf{V}$ cancel leaving the equation satisfied by $\hat{\mathbf{R}}$. This will be true for the denominator of the generalized likelihood ratio test as well. If $\hat{\mathbf{R}}_{|H_0}$ is a maximum-likelihood estimate of the covariance matrix under H_0, after a change of coordinates, then $\mathbf{V} \hat{\mathbf{R}}_{|H_0} \mathbf{V}^\dagger$ will be a maximum-likelihood estimate. Inserting these estimates into the generalized likelihood ratio formula leaves the arguments of the exponentials unchanged. The determinants may be factored

$$|\mathbf{V} \hat{\mathbf{R}} \mathbf{V}^\dagger| = |\mathbf{V}| |\hat{\mathbf{R}}| |\mathbf{V}^\dagger| \tag{218}$$

and canceled between the numerator and denominator, yielding the result that the test statistic is invariant with respect to the transformation of coordinates. One such transformation of coordinates is a change of scale; therefore, the test statistic is invariant with respect to a change in scale.

Using the above proof, it can also be shown that the transformation of coordinates for the constrained adaptive matched filter will be canceled in the test statistic, and the test statistic is invariant with respect to the transformation of coordinates.

Invariance to scale is a known property of the generalized likelihood ratio test [52]. This invariance to a change in coordinates does not show that the probability of false alarm will be independent of the unknown covariance. With invariance to the scale of the direction vector, it shows that the tests are invariant to the scale but not necessarily the structure of the true covariance matrix.

Independence of the density function of the test statistic to the structure of the true covariance matrix is an open research question. There is an interesting property of the test statistics that, if it were possible to condition on the estimated covariance matrix, would indicate independence of the test statistics to the structure of the true covariance matrix. This property holds when the following conditions are met.

- The sample covariance is not a maximum-likelihood estimate of the covariance. Since the dimension of the constraint space is less than the dimension of the space of possible sample covariance matrices, this typically holds with probability 1.

- The estimate of the constrained covariance matrix is within the interior of the constraint region. When this is true, then the necessary condition for the maximizer of the likelihood of Equation (213) is satisfied with equality.

- The possible variations of the covariance matrix must include the outer product of the signal direction vector.

When the above conditions are met, then Equation (216) is satisfied, and variations of the form $\delta\sigma\mathbf{d}\mathbf{d}^\dagger$ can be written

$$\frac{1}{P}\mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{d} = \mathbf{d}^\dagger\mathbf{R}^{-1}\mathbf{d} \quad . \tag{219}$$

The term on the right appears in both of the tests in the argument of the exponent of the likelihood equation under H_1 and in the denominator of the constrained adaptive matched filter. Substituting for the denominator of the constrained adaptive matched-filter test yields

$$\frac{\mathbf{d}^\dagger\hat{\mathbf{R}}^{-1}\mathbf{z}\mathbf{z}^\dagger\hat{\mathbf{R}}^{-1}\mathbf{d}}{\mathbf{d}^\dagger\hat{\mathbf{R}}^{-1}\mathbf{S}\hat{\mathbf{R}}^{-1}\mathbf{d}} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\gamma}{P} \quad . \tag{220}$$

98

If this test could be conditioned on the estimate of the covariance matrix, it would be easy to show that the resulting test statistic is independent of the true and estimated covariance matrices. As it is here, it serves to show that the same transformation on the primary and secondary data will be used to obtain the test statistic. This results in a comparison of the power of the primary data in the direction $\hat{\mathbf{R}}^{-1}\mathbf{d}$ to the average power of the secondary data in the same direction.

In the next chapter the methods presented in this chapter will be compared to the beamforming and detection methods that do not use structured covariance approaches.

9. ADAPTIVE BEAMFORMING AND DETECTION RESULTS

9.1 Introduction

The beamforming and detection methods that have been proposed in this report and some of the methods discussed in Chapter 3 will be compared in this chapter. It will be shown that using the knowledge that the true covariance matrix is structured, in the derivation of the signal-processing algorithms, can lead to a dramatic increase in performance.

An analytic expression has been found for only one of the structured covariance methods (Appendix A). In order to show that structured covariance methods may be applied to adaptive beamforming and detection, the methods for which analytic expressions have not been found will be simulated. These simulations will be used to compare the performance of the adaptive signal-processing algorithms in order to provide insight into how these methods would behave.

The parameters that may be varied for the simulations are common to both the beamformers and the detector methods. The parameters that may be varied are

- the number of sensors N,
- the sensor locations,
- the interference environment,
- the number of zero mean data vectors K,
- the receiver noise level $\mathrm{diag}(\sigma_1^2 \ldots \sigma_N^2)$,
- the number of terms in the finite sum approximation M,
- whether the constrained approaches are given knowledge of the receiver noise level and structure, and
- the direction corresponding to the mean term.

Rather than attempt to illustrate the performance for all combinations of the above parameters, the simulations will be restricted to two arrays and a fixed number of terms in the finite sum. The number of zero mean data vectors will also be fixed at $K = N$. The primary interest in these simulations is to show how the beamforming and detection methods perform in different noise environments.

The first array that will be simulated is a minimum-redundancy linear array with element spacings

$$\text{Array1} = 0, 0.5, 2.0, 3.0 \qquad . \qquad (221)$$

This vector indicates the location of the array elements in wavelengths along the array axis. The covariance matrix for this array has elements corresponding to six different non-zero "lags." The dimension of the constraint space is 13. The dimension of the entire space of 4 by 4 complex covariance matrices is 16. The only repeated covariance elements are along the diagonal of the constrained covariance matrices. The number of terms in the finite sum approximation was fixed at twice the dimension of the constraint space, and a diagonal term was estimated independently for both constrained methods. The resolution of the variable basis estimator was set to 0.1 degree.

The second array that is simulated is a thinned linear array or a uniform linear array with elements removed. The element spacings are

$$\mathrm{Array2} = [0, 0.5, 1.5, 2.0, 3.0, 3.5, 4.5, 5.0] \qquad (222)$$

The dimension of the constraint space in this case is 21. The resolution of the variable basis estimator was set to 0.05 degree, and the number of terms in the fixed set of bases is 42.

These arrays are linear arrays and cannot differentiate among sources that are along vectors rotated around the axis of the array; there is a "cone of ambiguity." Searching for directions and bases that maximize the likelihood then reduces from the two-dimensional search to a one-dimensional search.

Several different interference environments are simulated. These environments consist of discrete sources plus a diffuse background interference. The diffuse background interference will add a diagonal term to the covariance matrix, as will the receiver noise. The power level and location of the discrete sources that have been simulated for the 4-element array are shown in Table 3. The interference locations for the first three environments are represented by the basis

TABLE 3

Interference Directions of Arrival and Intensity; 4-Element Array

Model	Int. Sources	θ	Power	θ	Power
A	0				
B	1	61.8	40		
C	2	-34.80	40	61.80	50
D	1	57.3	60		

vectors that are used in the method of Chapter 6 but not by the initial bases for the method of Chapter 7. Environment D is not within the polyhedral cone formed by the fixed-bases estimator and is not in the initial cone formed by the variable bases estimator. For the 8-element array, slightly different environments were simulated, and these are shown in Table 4. The different

TABLE 4

Interference Directions of Arrival and Intensity; 8-Element Array

Model	Int. Sources	θ	Power	θ	Power	θ	Power
A	0						
B	2	58.75	40	-47.5	50		
C	6	-68.75	40	-34.75	30	20.5	30
		37.5	40	41.75	40	46.0	40

interference environments were used to provide very different noise covariance matrices so that the false alarm rate properties of the detectors can be investigated. The number of iterations of the iterative algorithms was fixed at 100.

9.2 Beamforming

The goal of the beamforming algorithms is to provide an estimate of the mean b while satisfying some optimality criterion. The criterion that will be used for comparison of the different beamforming methods is the output signal-to-noise ratio. Four different methods will be compared in this section. A summary of these methods is given below.

- The known noise covariance maximum-likelihood estimate of the mean. This method is an efficient estimator of the mean. The curves for this method will be labeled ML.

- The unknown noise covariance maximum-likelihood estimate of the mean. This method is equivalent to estimating the unstructured covariance matrix using the zero mean data samples and then using this estimate in the known noise covariance maximum-likelihood estimator. The curves for this method will be labeled AML.

- Estimate a structured covariance matrix satisfying the constraints imposed by the array geometry and operating environment, using only the data that is known to be

103

zero mean. Use this estimate in the known covariance maximum-likelihood estimator of the mean. The curves here will be labeled CML6 and CML7 for the estimation techniques of Chapters 6 and 7, respectively.

- Estimate jointly the structured covariance matrix and the mean using all of the data in a single estimation algorithm. The curves here will be labeled J6 and J7 for the methods of Chapters 6 and 7, respectively.

The covariance matrix estimation procedure based on the one-dimensional constraint space will not be discussed as a beamforming method. The signal-to-noise ratio will not exhibit a loss for this method since, assuming that the structure is correct, the signal-to-noise ratio is independent of the estimated covariance matrix. This is discussed in Appendix A.

9.2.1 Spatial Response

The first three beamforming methods provide estimates of the mean by a linear filtering operation on the non-zero mean data vector. Comparing the spatial response of these beamformers provides insight into how well the covariance estimate has been estimated. The known covariance maximum-signal-to-noise-ratio linear filter response is shown in Figure 20 for environment C. The responses for six realizations of the linear filter using the unconstrained maximum-likelihood covariance estimate are shown in Figure 21. Figures 22 and 23 show six realizations of the filter using constrained maximum-likelihood covariance estimate of Chapters 6 and 7, respectively. It should be emphasized that the sidelobes do not reflect the loss in signal-to-noise ratio that might be associated with a non-adaptive system for the same level of sidelobes. We can get a sense of the loss in signal-to-noise ratio by comparing the adaptive beamformers to the optimal; however, high sidelobes in themselves do not indicate poor performance.

Figures 24, 25, 26, and 27 show the beamformer response for the 8-element array and interference environment C where the difference in the beamformer responses is more obvious. In the next section the loss factor corresponding to these beamformer responses will be shown.

9.2.2 Signal-to-Noise Ratio Loss Factor

In this section the loss in signal-to-noise ratio that would occur through the use of the covariance estimates rather than the true covariance matrix in a linear beamformer will be compared. The density of this loss is known for the unconstrained covariance estimates, so the histogram of 200 realizations of the constrained covariance methods will be compared to this density function. Figures 28 and 29 are histograms of the loss factor from the simulations of the last two methods overlaid with the beta density function, which is the loss factor density when the beamformer is based on the unstructured covariance estimate. The histograms show that the constrained covariance estimates exhibit lower loss than the beamformer based on the unconstrained estimates. In these histograms, the true covariance was in the polyhedral cone formed by the finite basis used to estimate the structured covariance.

104

beamformer response
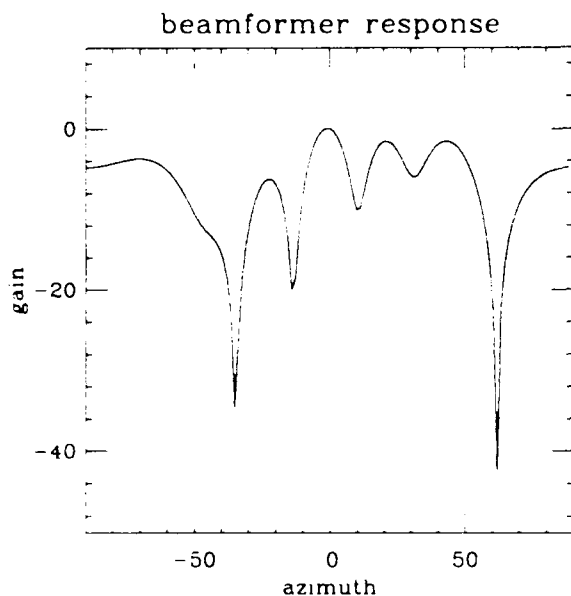
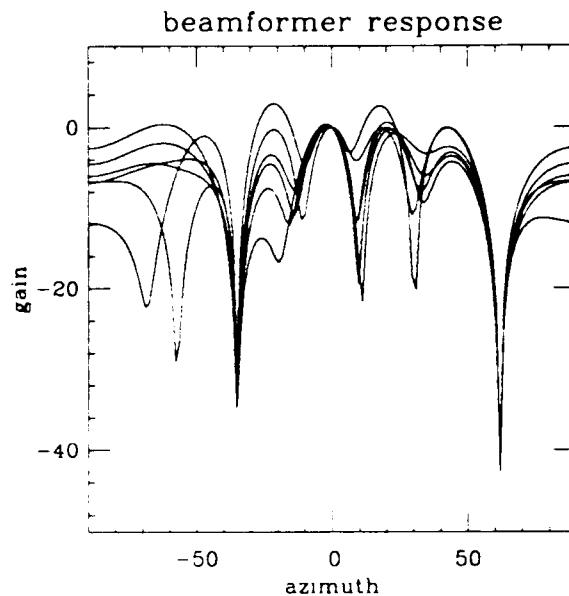Figure 20. Beamformer response using known covariance.

beamformer response

Figure 21. Beamformer response using unstructured covariance estimate.
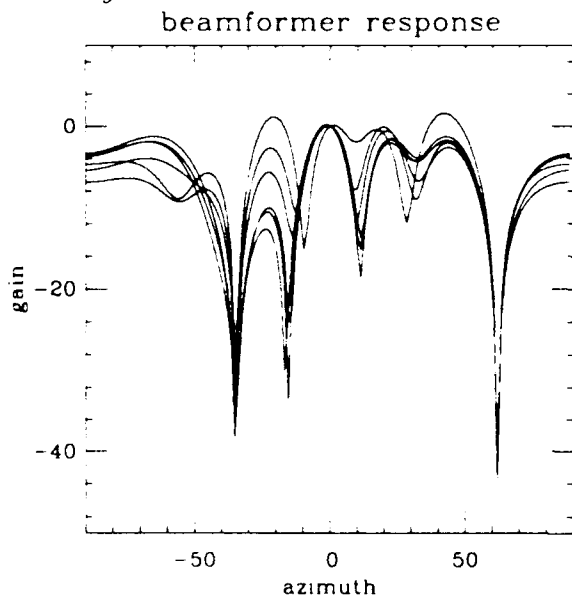
beamformer response

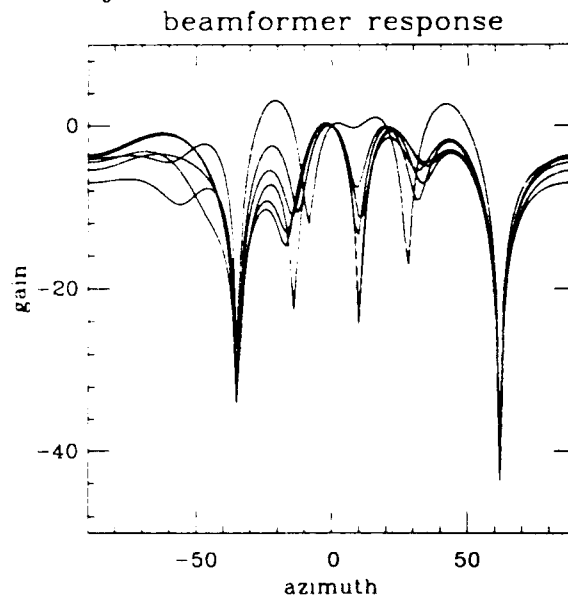Figure 22. Beamformer response using Chapter 6 structured covariance estimate.

beamformer response

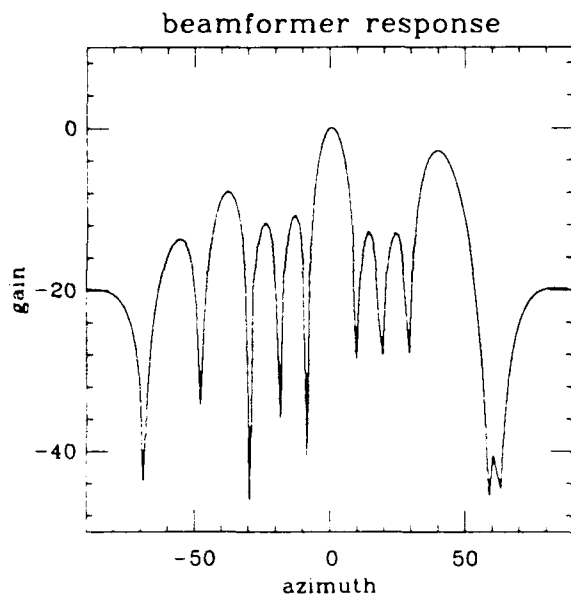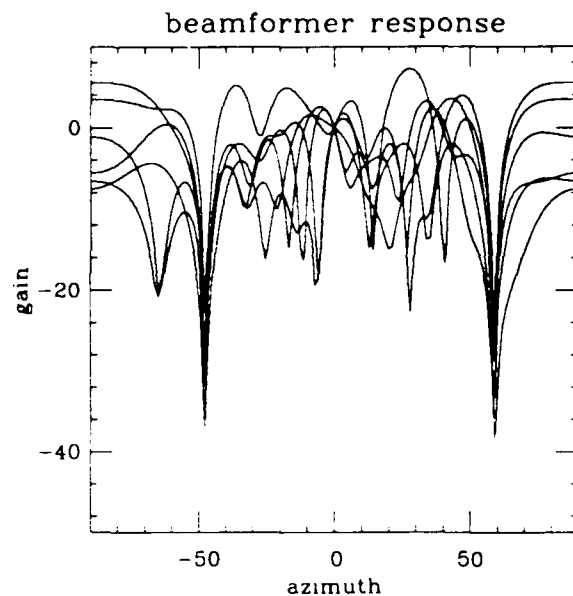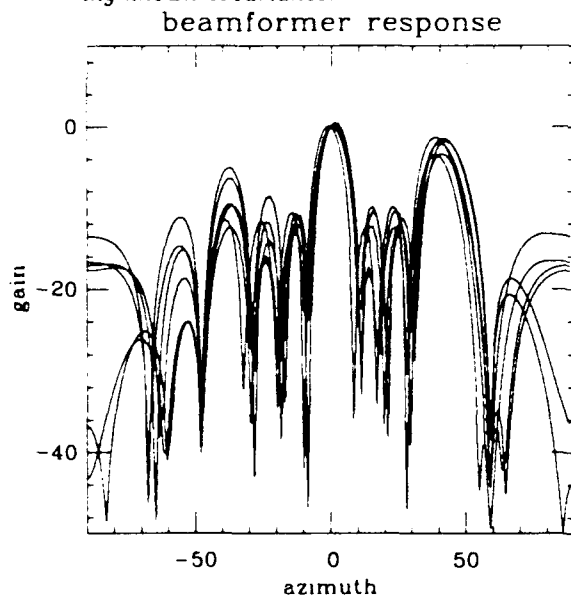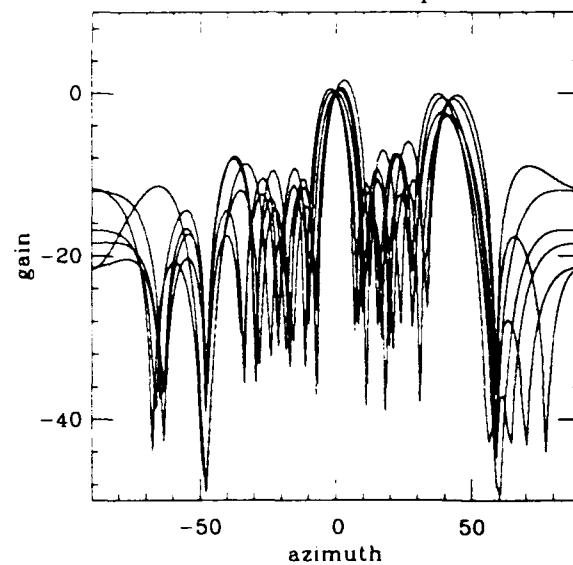Figure 23. Beamformer response using Chapter 7 structured covariance estimate.

105

beamformer response

beamformer response

Figure 24. Beamformer response using known covariance.

Figure 25. Beamformer response using unstructured covariance estimate.

beamformer response

beamformer response

Figure 26. Beamformer response using Chapter 6 structured covariance estimate.

Figure 27. Beamformer response using Chapter 7 structured covariance estimate.

In Figures 30 and 31, the true covariance matrix is not within the polyhedral cone formed by the fixed basis. There is a single interference source that is located at 57.3 degrees with an intensity of 60 dB more than the receiver noise (environment D). The fixed-basis estimator of Chapter 6 shows a signal-to-noise ratio loss, which is much greater than the loss for the estimator whose bases are allowed to vary.

These figures show that there can be an increase in the signal-to-noise ratio if there is an adequate number of terms in the finite sum used to form the covariance matrix when using the method of Chapter 6; however, when there is an insufficient number of terms, then this method may perform more poorly than the unconstrained approach. There is virtually no change in the loss for the method of Chapter 7 for the two different interference environments. This shows that the method of Chapter 7 will provide covariance estimates that are closer to the true covariance matrix for this case.

9.2.3 Bias and Variance

In this section, the bias and variance of the estimators for the 4-element array are discussed. It is known that all of the beamforming methods, except the joint estimation of the mean and covariance, yield unbiased estimators of the mean. In this section, simulations are used to compare the bias of this method to the bias of the other methods. In this simulation, 200 realizations of the secondaries are used to form weight vectors for the unconstrained and both constrained approaches; then, an additional vector realization of the same process is generated. Different signal amplitudes corresponding to a desired signal-to-noise ratio are added to this vector. The mean is estimated using the weight vectors already formed as well as by jointly estimating the mean with the structured covariance matrix using the estimation methods of Chapters 6 and 7. The true mean and the estimates are saved to a file where the sample statistics are formed.

The covariances that are used in these simulations are within the polyhedral cone formed by the finite basis estimator of Chapter 6. The point at -20 dB signal-to-noise ratio is actually for no signal so that the bias and variance of the estimators with no input can be shown. The variance for the known covariance matched filter is 1, and the variance of the estimate that uses the unconstrained covariance estimate is 2.5. Two interference environments were simulated: the first is environment A, and the second is environment C.

It can be seen from Figures 32 and 33 that the variance of the estimates of the mean using the structured methods appears to be uniformly better than those where the structure of the covariance is not taken into account. All of these methods except the one that jointly estimates the mean and the covariance are known to be unbiased. Both of the estimates that jointly estimate the mean and the covariance show a variation in the bias with changes in the signal-to-noise ratio. Figures 34 and 35 are based on the interference environment C. It is not clear from these simulations or others that are not shown if there is a trend in the bias or variance of the estimate of the mean. The true mean varies from 1 to 20 for the signal-to-noise ratio varying from 0 to 26 dB.
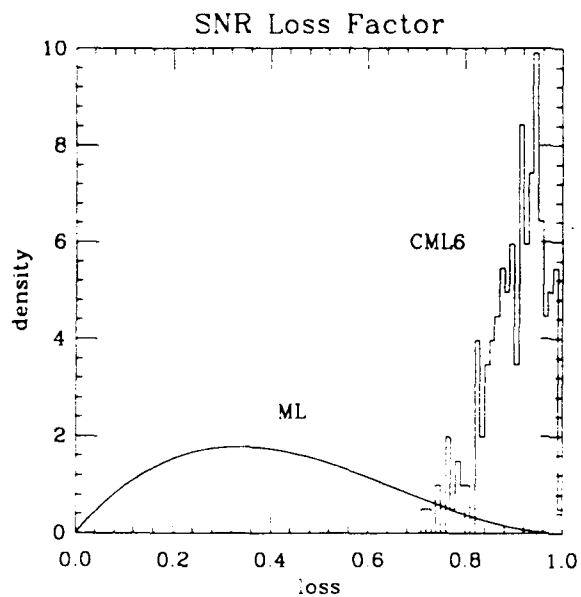
SNR Loss Factor

CML6

ML

loss

Figure 28. Signal-to-noise ratio loss factor. Chapter 6 and unconstrained methods.

SNR Loss Factor

CML7

ML
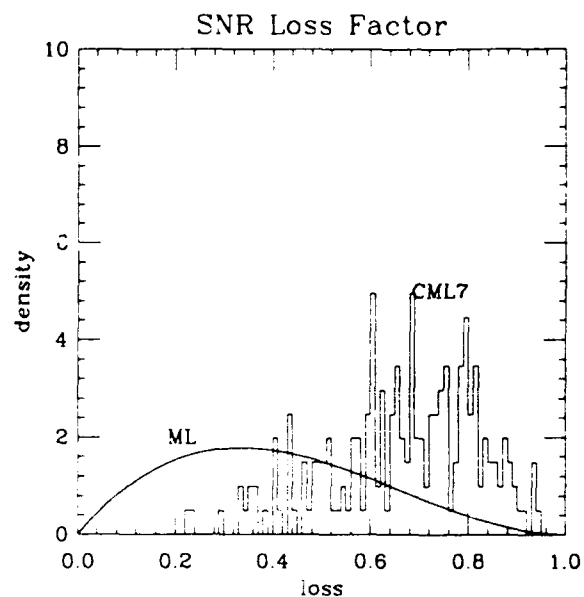
loss

Figure 29. Signal-to-noise ratio loss factor. Chapter 7 and unconstrained methods.

SNR Loss Factor

CML6

ML
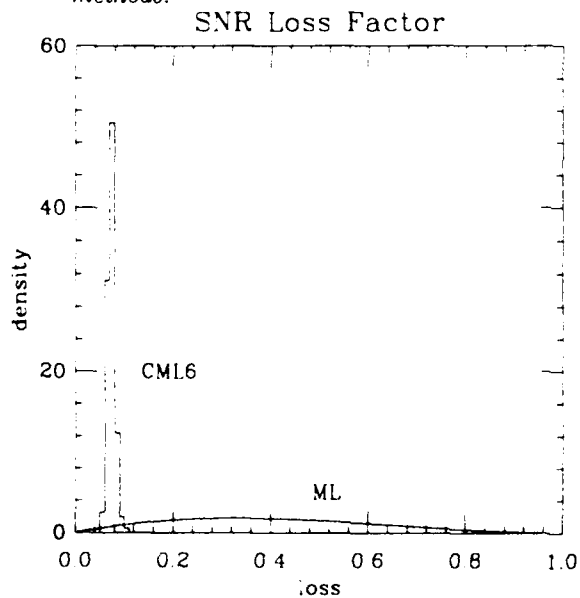
loss

Figure 30. Signal-to-noise ratio loss factor. Chapter 6 and unconstrained methods.

SNR Loss Factor

CML7

ML
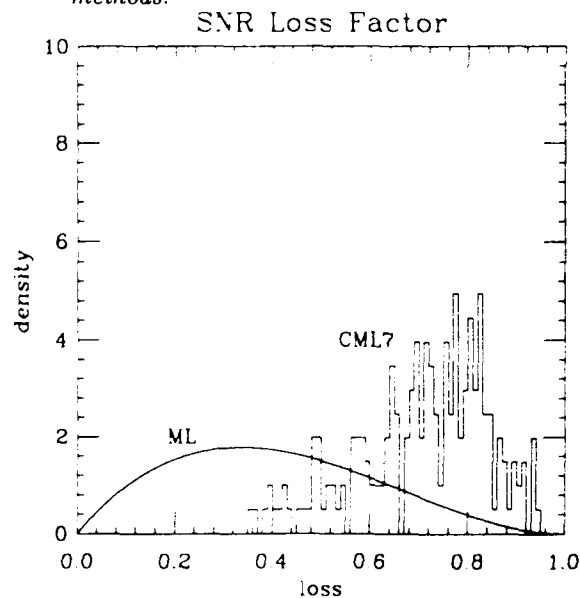
loss

Figure 31. Signal-to-noise ratio loss factor. Chapter 7 and unconstrained methods.
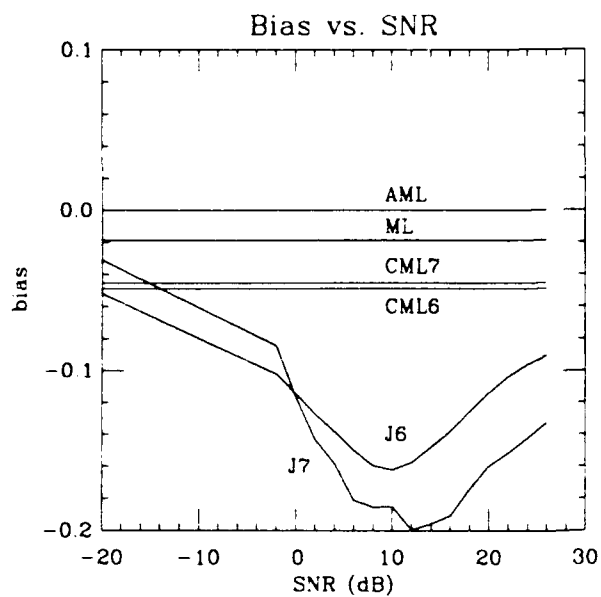
Figure 32. Bias of the mean estimates.
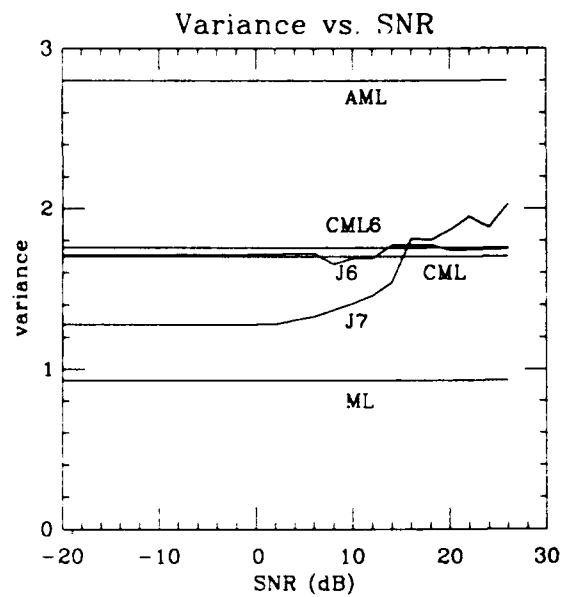
Figure 33. Variance of the mean estimates.
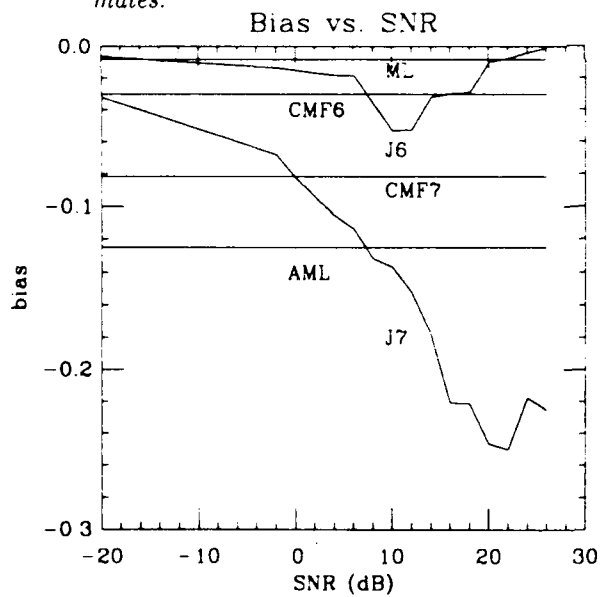
Figure 34. Bias of the mean estimates.
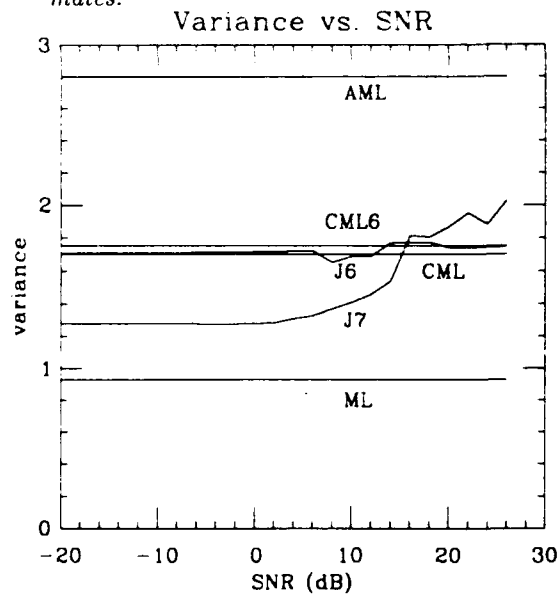
Figure 35. Variance of the mean estimates.

109

9.3 Adaptive Detection

There have been many different detectors discussed in this report. These detectors, the probability of false alarm, and the probability of detection are shown in Table 5. Of these detectors

TABLE 5

Detector Performance Comparisons

Detector	Abbreviation	Know	Constrained?
Generalized LR	MF	R	-
Generalized LR	GLR	-	N
Adaptive MF	AMF	-	N
Generalized LR	SIGI	R*	Y
Generalized LR	CGLR	-	Y
Adaptive MF	CAMF	-	Y
* Known to a scale factor			

the first four have analytic expressions describing the performance. The last two must be simulated. The last four of these detectors have been proposed or analyzed as a portion of this report. The false alarm and detection probabilities will be estimated by the use of Monte Carlo simulations.

Since simulations are used to illustrate the performance of the detection methods, the probabilities of false alarm and detection will be restricted to what would be considered fairly high levels so that the estimates of these probabilities are reliable. The necessity of keeping the probabilities at high levels can be shown by examining the variance of the estimates of the probabilities. Defining the indicator function as

$$I(x) = \begin{cases} 1 & x \geq \alpha \\ 0 & x < \alpha \end{cases} \quad , \tag{223}$$

then for a single sample and for the test statistic density function $f(x)$,

$$E\{I(x)\} = \int_{\alpha}^{\infty} f(x)\,dx = PFA \quad , \tag{224}$$

110

and

$$E\{(I(x))^2\} = \int_\alpha^\infty f(x)\,dx = PFA \qquad ; \tag{225}$$

then the variance is

$$\sigma^2 = PFA - PFA^2 \qquad . \tag{226}$$

For J realizations, thresholding the test statistic and averaging the number of occurrences where the test statistic is greater than the threshold will give an unbiased estimate of the probability of detection or probability of false alarm. The variance of the estimate for the average of J realizations is $\sigma^2 = (1/J)(PFA - PFA^2)$. From this it can be seen that unless a large number of samples is used to form an estimate of the false alarm or detection probability, the variance will be high. If Tchebycheff's inequality is used to indicate the probability that the estimate deviates from the true probability, the result for 200 realizations is

$$P(|x - PFA| \geq a * PFA) \leq \frac{\sigma^2}{a^2 PFA^2} = \frac{1}{200a^2}\left[\frac{1}{PFA} - 1\right] \qquad . \tag{227}$$

Using this equation, for a probability of detection or false alarm of 0.1, the probability that any realization varies by more than 50 percent from this probability is 0.18.

The estimates of the detection probabilities will not be entirely independent since for the 4-element array only 200 covariance estimates will be performed. These 200 estimates are used to calculate the test statistic for 16 independent realizations of the primary vector at 16 different signal-to-noise ratios for each covariance estimate. For the full constrained generalized likelihood ratio test, 200 realizations of the secondary data will be used along with the 16 independent realizations of the primary vector for each of the realizations of the secondary data. This yields 3200 realizations of the test statistics for each of the detection methods. For the 8-element array 100 realizations of the secondary data are used rather than 200. One hundred iterations of the iterative methods are used to form the estimates. For the joint estimates of the mean and the covariance, the estimate of the covariance matrix under H_0 was formed; then this estimate was used to initialize the algorithms for H_1.

First, the two detectors for which a closed-form expression for false alarm probabilities have not been found are simulated. The estimated false alarm probabilities versus the threshold are shown in Figures 36 and 37 for the constrained adaptive matched-filter detector and for environments A through C for the 4-element array. Figures 38 and 39 illustrate the false alarm probabilities for the constrained generalized likelihood ratio detectors. The test statistic that is used here for the likelihood ratio tests is the log of the $K + 1$ root of the ratio of the likelihoods found by the methods of Chapters 6 and 7.

The probabilities of false alarm versus threshold for the 8-element array are shown in Figures 40, 41, 42, and 43. It should be noted that the probability of false alarm is very sensitive to changes in the threshold or the density of the test statistic.

Before discussing the simulation results, the performance results for the detectors with closed-form expressions for the detection probabilities are shown. In Figures 44 and 45 the probability of detection for the first four detectors of Table 5 is shown. It can be seen that knowing the noise statistics adds appreciably to the detector performance when the detectors are compared to the approaches that are not given any knowledge of the noise statistics.

The probabilities of detection for the detectors introduced in this report will now be compared to the other methods. In order to plot the probability of detection, a desired false alarm probability was chosen. The probabilities of false alarm for the three realizations of the simulated methods are averaged, and a threshold is selected for each detector and each estimation method. The desired probability of false alarm was 0.1; the probabilities of false alarm for the realizations varied from 0.07 to 0.135 for the 4-element array and 0.06 to 0.17 for the 8-element array. The probability of detection for the methods where an analytic expression exists was calculated for the desired false alarm probability. The probabilities of detection are shown for the three interference environments discussed earlier for the two simulated arrays where the true covariance is within the polyhedral cone.

Figures 46 through 49 show the simulation results for the 4-element array, and Figures 50 through 53 show the results for the 8-element array. It can be seen that the probabilities of detection at a particular probability of false alarm show a dramatic improvement compared to the unconstrained adaptive detection methods. The probability of detection for the detectors using the constrained covariance is nearly that of the known covariance matched-filter detector. The additional signal-to-noise ratio that would be required to achieve the same probability of detection as the MF detector is less than 5 dB. Due to the statistical variation, there are some realizations that estimate a probability of detection that would be higher than that of the known covariance test.

In the detector simulations, the method of Chapter 6 appears to provide detectors that have a higher probability of detection compared to those that use the estimation method of Chapter 7. This is due to the location of the interference sources with respect to the directions that parameterize the bases. Since there is a relatively small number of bases, $(2L)$, the volume of the constraint space is smaller for this method, and the truth is interior to this smaller volume. The estimator of Chapter 6 effectively has more knowledge of the interference environment than the estimator of Chapter 7. When the true covariance matrix is structured, but not in the polyhedral cone formed by the smaller number of bases, then detection performance can suffer. Figure 54 illustrates this for the interference of environment D. The thresholds used to generate these curves were those determined by the other simulations. The curves are labeled CAMF6 and CAMF7 for the constrained adaptive matched-filter detector and are labeled CLR6 and CLR7 for the constrained likelihood ratio tests. The probability of detection does not vary appreciably when the variable bases estimator is used
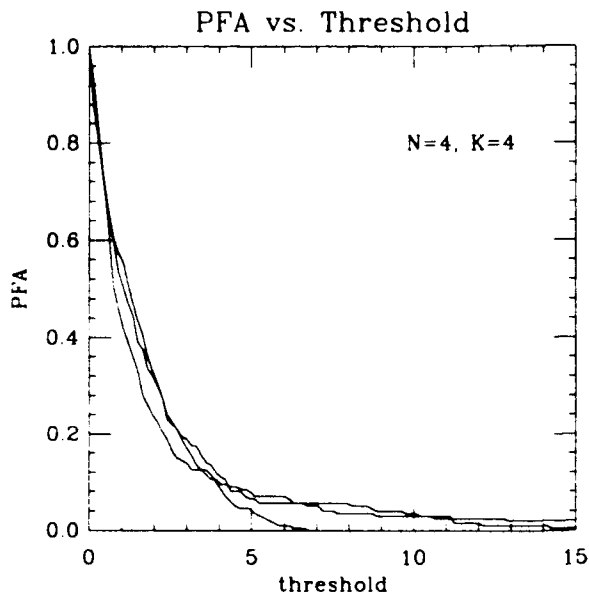
PFA vs. Threshold

Figure 36. Probability of false alarm vs threshold. Chapter 6 constrained AMF.
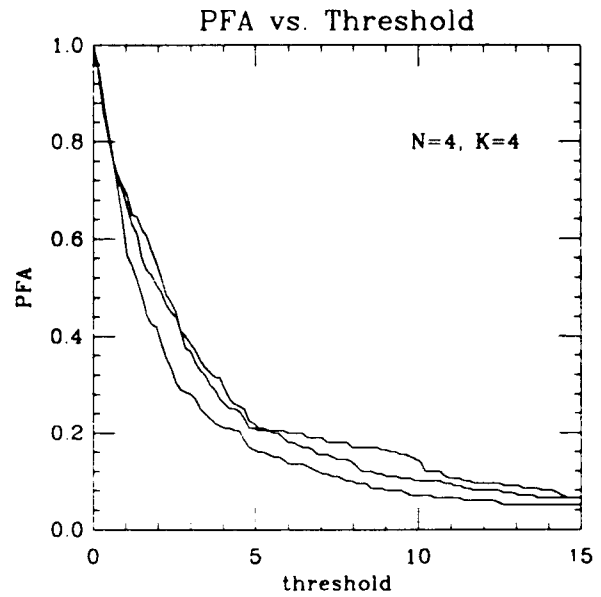
PFA vs. Threshold

Figure 37. Probability of false alarm vs threshold. Chapter 7 constrained AMF.
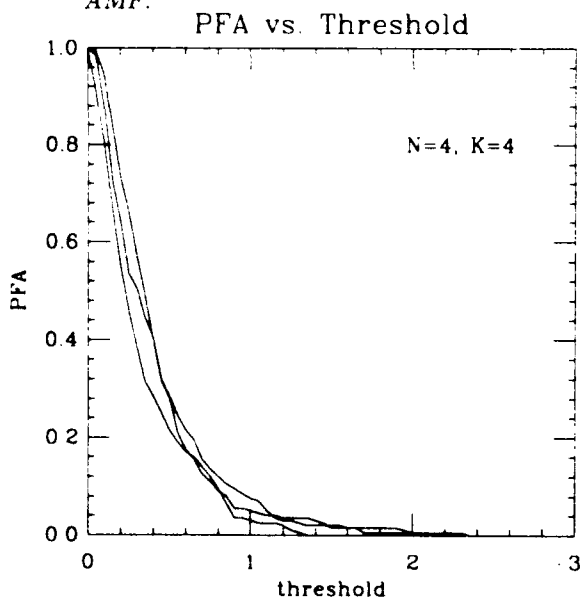
PFA vs. Threshold

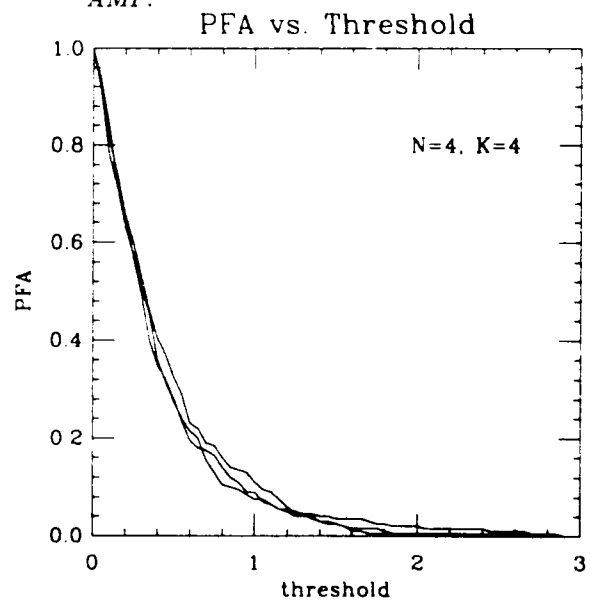Figure 38. Probability of false alarm vs threshold. Chapter 6 likelihood ratio.

PFA vs. Threshold

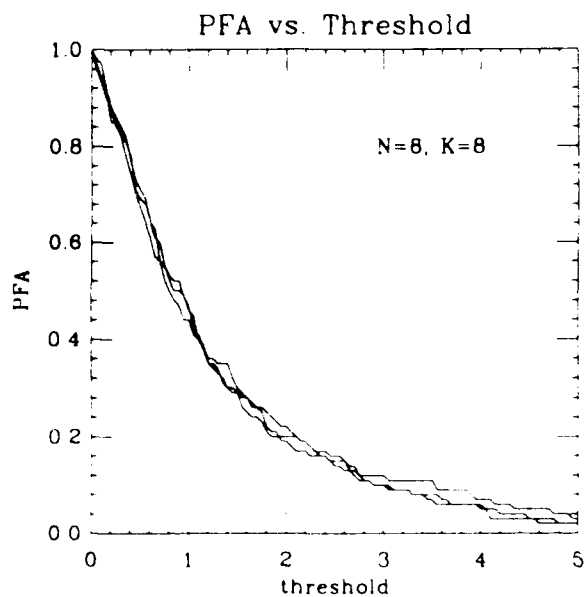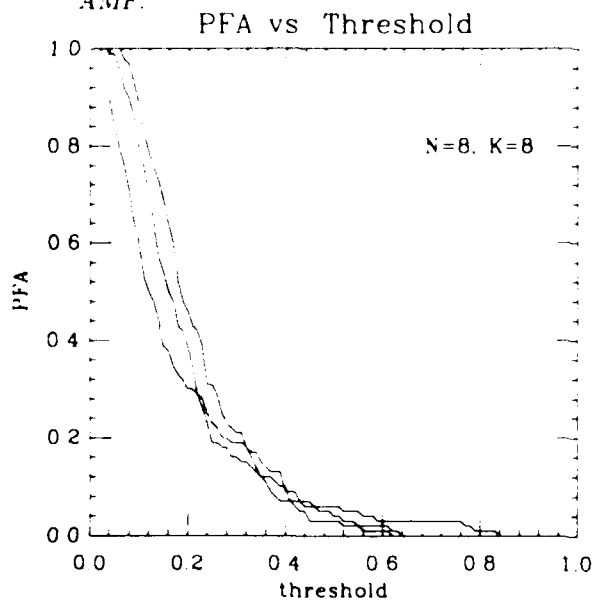Figure 39. Probability of false alarm vs threshold. Chapter 7 likelihood ratio.

113

PFA vs. Threshold

Figure 40. Probability of false alarm vs threshold. Chapter 6 constrained AMF.
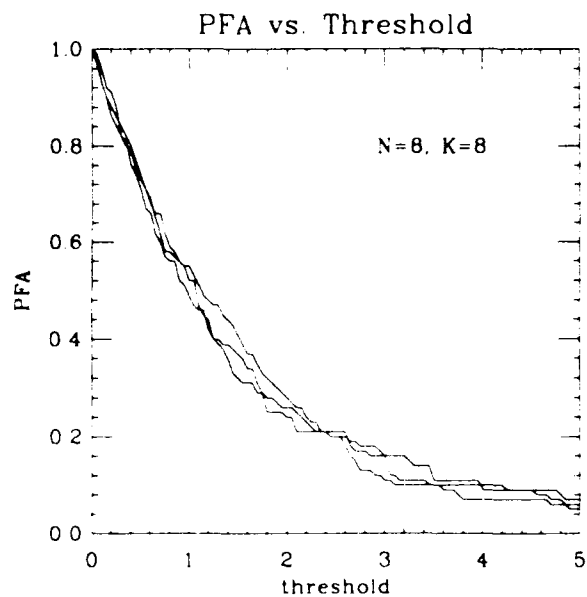
PFA vs. Threshold

Figure 41. Probability of false alarm vs threshold. Chapter 7 constrained AMF.

PFA vs Threshold

Figure 42. Probability of false alarm vs threshold. Chapter 6 likelihood ratio.
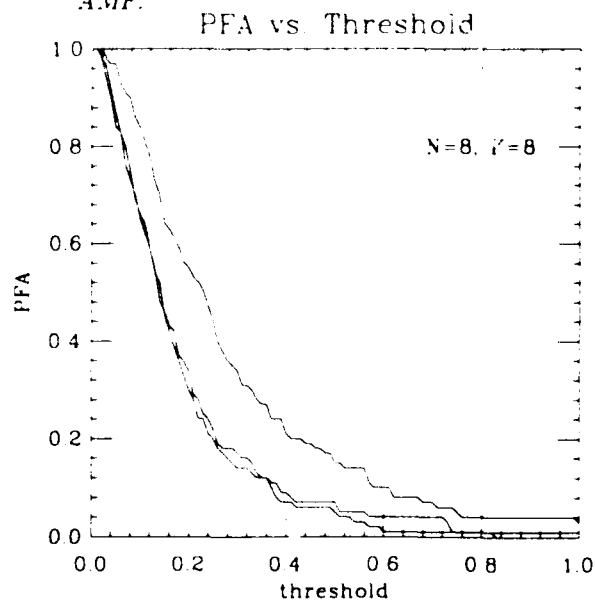
PFA vs. Threshold

Figure 43. Probability of false alarm vs threshold. Chapter 7 likelihood ratio.

114
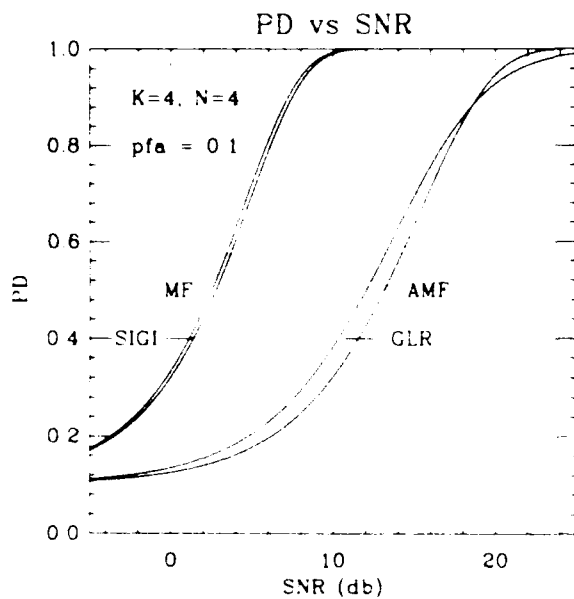
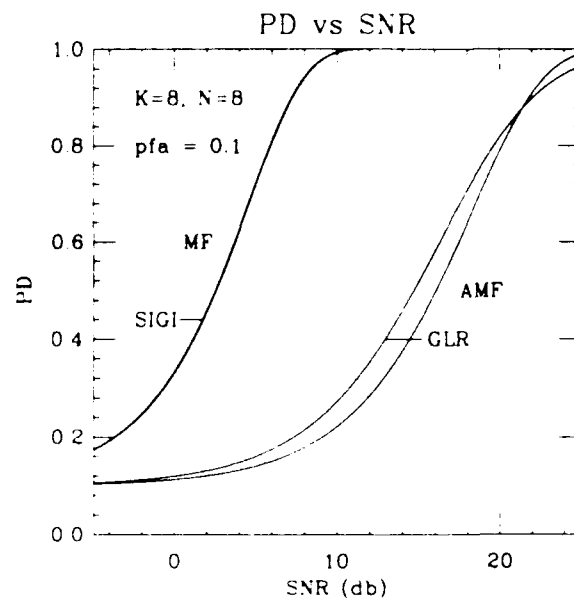Figure 44. Probability of detection using analytical expressions.

Figure 45. Probability of detection using analytical expressions.

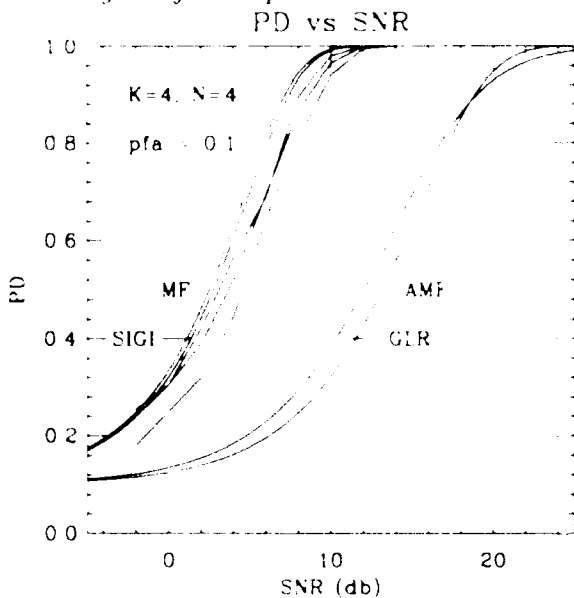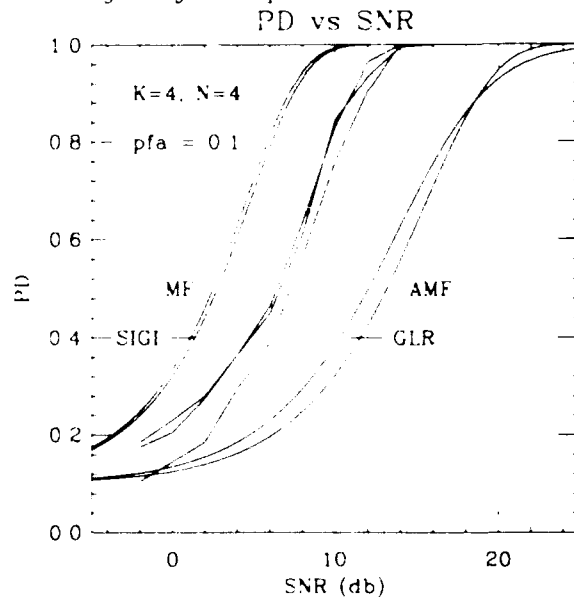Figure 46 Probability of detection from simulations. Chapter 6 constrained AMF

Figure 47. Probability of detection from simulations. Chapter 7 constrained AMF.

to generate the test statistics: although, there is 10 dB or more additional loss for the fixed-bases methods.

9.4 Conclusion

In this chapter, the adaptive signal processing methods proposed in this report have been compared. The simulations show that there can be a dramatic improvement in performance by making use of the complete data model.
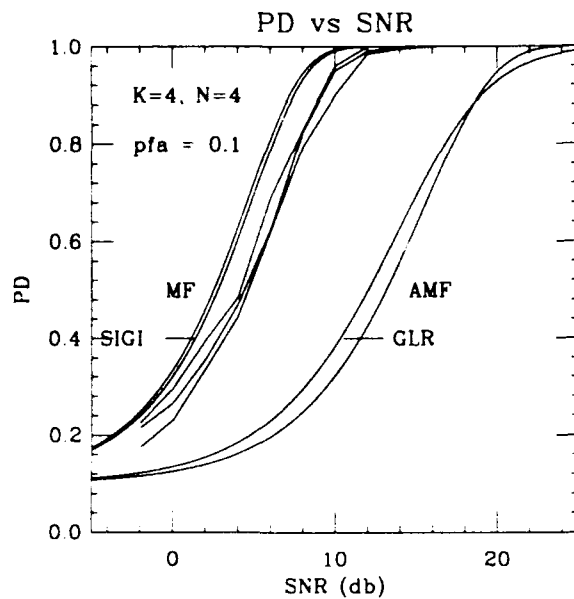
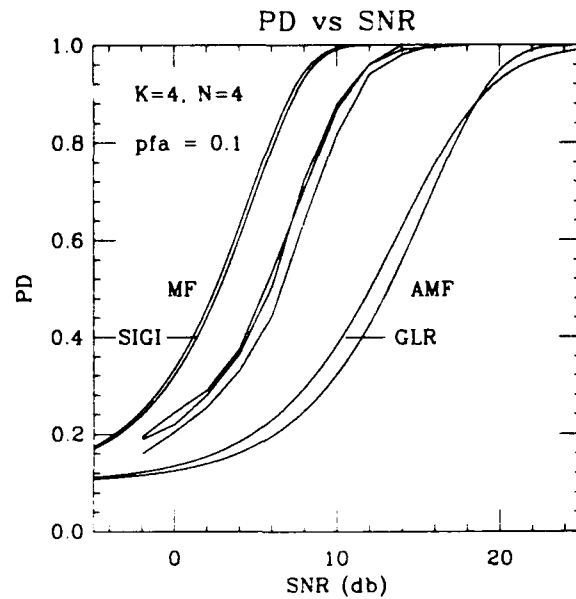*Figure 48. Probability of detection from simulations. Chapter 6 likelihood ratio.*

Figure 49. Probability of detection from simulations. Chapter 7 likelihood ratio.
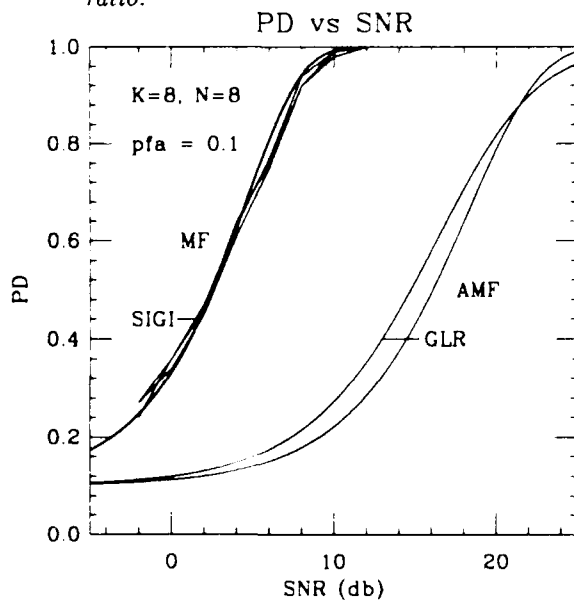

Figure 50. Probability of detection from simulations. Chapter 6 constrained AMF.
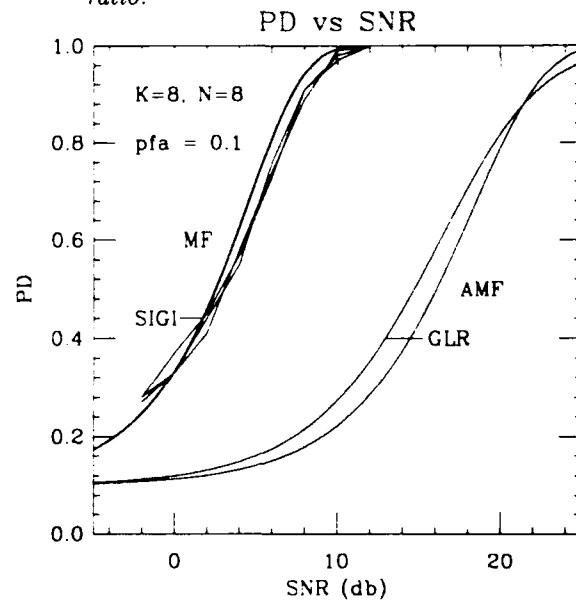
Figure 51. Probability of detection from simulations. Chapter 7 constrained AMF.

117

PD vs SNR

K=8, N=8

pfa = 0.1

MF

SIGI

AMF

GLR

PD

SNR (db)
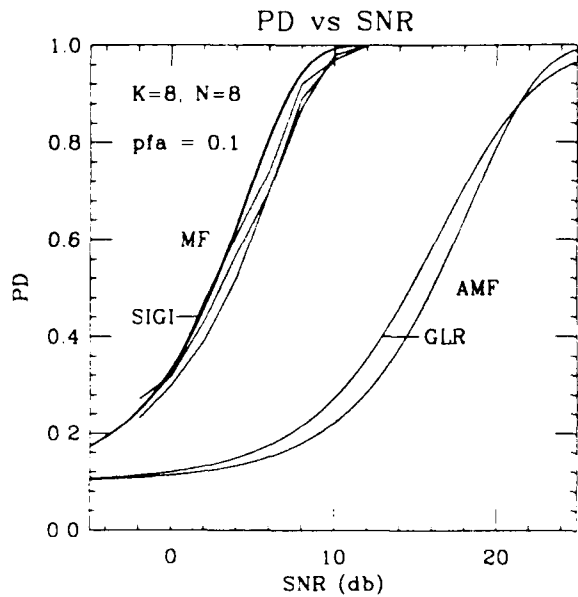
Figure 52. Probability of detection from simulations. Chapter 6 likelihood ratio.

PD vs SNR

K=8, N=8

pfa = 0.1

MF

SIGI

AMF

GLR

PD

SNR (db)
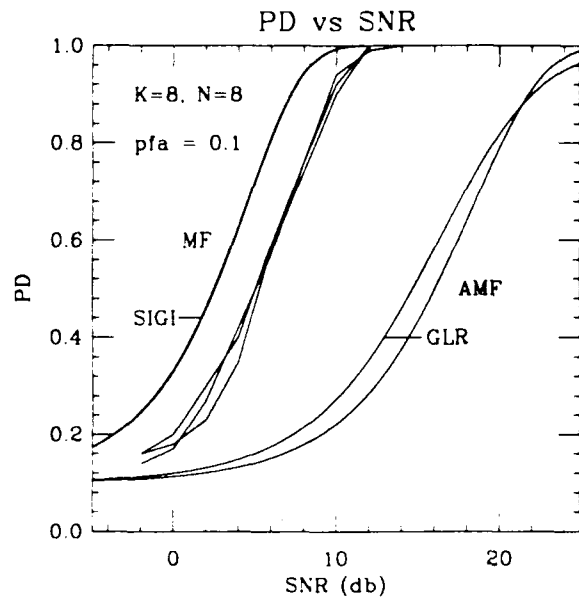
Figure 53. Probability of detection from simulations. Chapter 7 likelihood ratio.

PD vs SNR

K=4, N=4

pfa = 0.1

CAMF7

CLR7

MF

GLR
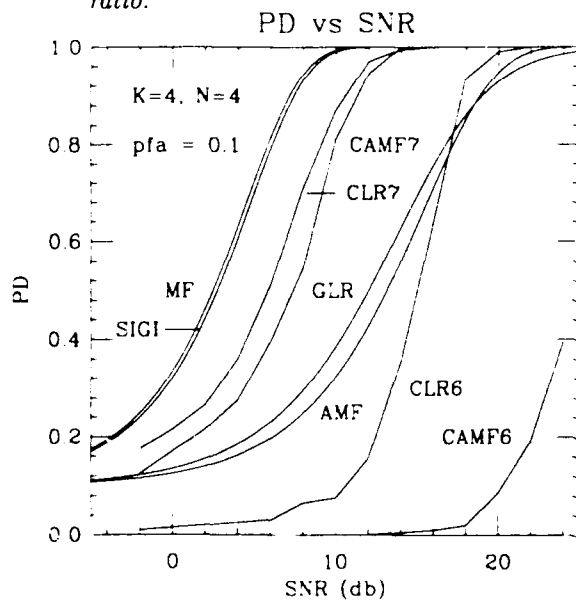
SIGI

CLR6

AMF

CAMF6

PD

SNR (db)

Figure 54. Probability of detection from simulations; Chapter 6 constrained AMF.

118

10. CONCLUSIONS

The subject of this report has been the use of sensor arrays to perform beamforming and detection. Maximum-likelihood techniques have been used to develop approaches which can be implemented to accomplish these tasks. The results of this report and possible future research areas are summarized below.

The radar detection and the communications applications of this report were discussed in Chapter 2, and the resulting data model was introduced. The radar detection model is the well-known non-fluctuating target model for which target presence is indicated by a non-zero structured mean. The structured mean is modeled as an unknown scalar times an array steering or direction vector. In this model, the covariance matrix is composed of two terms. The first is a diagonal matrix due to noise sources internal to the receivers; the second is generated by an integral expression consisting of a positive spatial-temporal power-spectral density and the outer products of the array direction vectors. The communications model is similar to the detection model, with the transmitted information content contained in the structured mean.

There are several modifications of the data model that would lead to areas of future research. One model could remove the assumption that the direction of a target and the data vector containing a possible target return are known. Multiple targets could then be allowed. As shown in Chapter 6, the likelihood can be unbounded for some conditions; increasing the number of sample vectors that can contain a mean will make this problem worse. A straightforward maximization of the likelihood may not be feasible, and methods of preventing this from occurring would need to be developed.

An additional model which could be the subject of further research would be to remove the assumption of independence for the data samples. A detector which could detect a moving target in clutter based on the returns from many samples could then be developed.

Chapter 3 is used to review some of the current methods of beamforming and detection. It was pointed out that optimal beamforming and detection require knowledge of the noise statistics. As these statistics are seldom known, adaptive approaches are utilized to estimate the statistics. If the amount of data is limited, then there can be an appreciable loss in performance for the adaptive approaches motivating the use of a more detailed data model.

In Chapter 4, a detector is derived and analyzed. The derivation uses the structured mean of the data model but does not use any knowledge of the interference environment or the array when estimating the noise covariance matrix. This detector provides a desirable simplification of the full generalized likelihood ratio test and provides similar performance for signals which are in alignment with the assumed signal direction. An original contribution discussed in this report is the analysis of detection performance when the signal is not aligned with the assumed signal direction.

The set of noise covariance matrices which is possible for a particular array under the model discussed in Chapter 2 is the subject of Chapter 5. The set of possible covariance matrices is

characterized as a convex cone that is a subset of an N^2 dimensional real vector space. Structured covariance matrices are defined to be a convex cone in a lower dimensional space. Some of the representations for covariance matrices that are structured are introduced in this chapter. These representations are introduced with the intent that they be used to enforce the structure on estimated covariance matrices that are members of the constraint space.

One of the representations introduced in Chapter 5 approximates the integral expression for the constraint space by a finite sum. The result consists of a sum of outer products of the array direction vectors weighted by samples of the spatial-temporal power-spectral-density function. A joint maximum-likelihood estimator of the structured mean and covariance matrix are developed in Chapter 6. The resulting iterative procedure is based on the Expectation-Maximization algorithm and results in a sequence of estimates for which the log-likelihood is non-decreasing. Stable points of the resulting estimator satisfy the necessary conditions for a maximizer of the likelihood function. Certain technical aspects of the resulting estimation algorithm are discussed in this chapter as well.

Properties of the likelihood surface and a maximum-likelihood estimate of the weight of single spectral point are discussed as an introduction to Chapter 7. Based on this discussion and one of the representations of Chapter 5, an algorithm for estimating constrained means and covariance matrices is introduced. This method does not require the large number of spectral weights but provides resolution and performance advantages compared to the estimator of Chapter 6. The likelihood for the sequence of estimates produced by this estimator is non-decreasing, and stable points of the resulting algorithm satisfy the necessary conditions for a maximizer of the likelihood.

Applying the structured estimators to beamforming and detection is the subject of Chapter 8. Two methods of utilizing the estimates produced in Chapters 6 and 7 to perform both beamforming and detection are discussed. Several properties of these methods are shown, including bias and some aspects of CFAR behavior.

The methods of beamforming and detection discussed in this report are compared in Chapter 9. It is shown that there can be a dramatic increase in signal-to-noise ratio for the adaptive beamformers based on the structured covariance estimates compared to beamformers which are not based on this knowledge. This is seen in the signal-to-noise ratio loss factor and in the beamformer responses. For the detectors which use knowledge of the array geometry and the noise environment, there is a similar increase in performance. The false alarm probability estimated by the simulations has a statistical variation that would be within what would be expected if the detectors had the CFAR property. The probability of detection shows the dramatic increase in performance that was seen in the constrained adaptive beamformers. The loss in signal-to-noise ratio for the same probability of detection is much less than the methods which do not utilize the knowledge of the covariance structure.

APPENDIX A
$\sigma^2 \mathbf{I}$ CONSTRAINED DETECTOR

A convenient test that may be used as a reference occurs when the covariance is constrained to be of the form $\sigma^2 \mathbf{I}$. When the covariance is known with the exception of a scale factor, the results of this appendix may be used by transforming the data to the $\sigma^2 \mathbf{I}$ case through whitening.

The notation used here is identical to that used in Chapter 4. A single primary vector $\mathbf{z}$ is zero mean on H_0 and has a mean of the form $b\mathbf{d}$ under H_1. In addition, K mutually independent zero mean secondary vectors, the columns of $\mathbf{Z}$, are available.

For a single vector sample with mean $b\mathbf{d}$, the complex Gaussian density function is

$$f(\mathbf{z}) = \pi^{-N} \sigma^{-2N} e^{\frac{-1}{\sigma^2}(\mathbf{z}-b\mathbf{d})^\dagger(\mathbf{z}-b\mathbf{d})} \qquad . \tag{A.1}$$

The generalized likelihood ratio test over the joint density function of $(\mathbf{z}, \mathbf{Z})$ is

$$\Lambda = \frac{\max\limits_{b,\sigma_1^2} f_{\mathbf{z},\mathbf{Z}|H_1}(\mathbf{z}, \mathbf{Z}; b, \sigma_1^2 | H_1)}{\max\limits_{\sigma_0^2} f_{\mathbf{z},\mathbf{Z}|H_0}(\mathbf{z}, \mathbf{Z}; \sigma_0^2 | H_0)} \begin{array}{c} H_1 \\ \gtrless \\ H_0 \end{array} \alpha \qquad . \tag{A.2}$$

Substituting in the density function and canceling common terms yields

$$\Lambda = \frac{\max\limits_{b,\sigma_1^2} \sigma_1^{-2N(K+1)} e^{\frac{-1}{\sigma_1^2}[(\mathbf{z}-b\mathbf{d})^\dagger(\mathbf{z}-b\mathbf{d})+\sum_{i=1}^{K}\mathbf{z}_i^\dagger\mathbf{z}_i]}}{\max\limits_{\sigma_0^2} \sigma_0^{-2N(K+1)} e^{\frac{-1}{\sigma_0^2}[(\mathbf{z}^\dagger\mathbf{z}+\sum_{i=1}^{K}\mathbf{z}_i^\dagger\mathbf{z}_i]}} \qquad . \tag{A.3}$$

The maximum likelihood estimate for σ_1^2 is found by maximizing the numerator with respect to σ_1^2 with the result that

$$\hat{\sigma}_1^2 = \frac{1}{N(K+1)}[(\mathbf{z}-b\mathbf{d})^\dagger(\mathbf{z}-b\mathbf{d}) + \sum_{i=1}^{K}\mathbf{z}_i^\dagger\mathbf{z}_i] \qquad , \text{and} \tag{A.4}$$

likewise

$$\hat{\sigma}_0^2 = \frac{1}{N(K+1)}[\mathbf{z}^\dagger\mathbf{z} + \sum_{i=1}^{K}\mathbf{z}_i^\dagger\mathbf{z}_i] \qquad . \tag{A.5}$$

Substituting these estimates into Equation (A.3) yields

$$\Lambda = \left(\frac{\frac{1}{N(K+1)}[(\mathbf{z} - b\mathbf{d})^{\dagger}(\mathbf{z} - b\mathbf{d}) + \sum_{i=1}^{K} \mathbf{z}_i^{\dagger}\mathbf{z}_i]^{-N}}{\frac{1}{N(K+1)}[\mathbf{z}^{\dagger}\mathbf{z} + \sum_{i=1}^{K} \mathbf{z}_i^{\dagger}\mathbf{z}_i]^{-N}} \right)^{K+1} \frac{e^{-N(K+1)}}{e^{-N(K+1)}} \quad . \quad (A.6)$$

Canceling common terms and taking the $N(K+1)$ root (a monotonic functional of the likelihood) yields

$$\Lambda = \frac{\max_b [(\mathbf{z} - b\mathbf{d})^{\dagger}(\mathbf{z} - b\mathbf{d}) + \sum_{i=1}^{K} \mathbf{z}_i^{\dagger}\mathbf{z}_i]^{-1}}{[\mathbf{z}^{\dagger}\mathbf{z} + \sum_{i=1}^{K} \mathbf{z}_i^{\dagger}\mathbf{z}_i]^{-1}} \mathop{\gtrless}_{H_0}^{H_1} \alpha \quad . \quad (A.7)$$

α has been redefined to include the effects of the $N(K+1)$ root.

Maximizing the numerator with respect to b is equivalent to minimizing

$$(\mathbf{z} - b\mathbf{d})^{\dagger}(\mathbf{z} - b\mathbf{d}) = \mathbf{z}^{\dagger}\mathbf{z} - \frac{|\mathbf{d}^{\dagger}\mathbf{z}|^2}{\mathbf{d}^{\dagger}\mathbf{d}} + \mathbf{d}^{\dagger}\mathbf{d}|\frac{\mathbf{d}^{\dagger}\mathbf{z}}{\mathbf{d}^{\dagger}\mathbf{d}} - b|^2 \quad , \quad (A.8)$$

which is clearly minimized by

$$\hat{b} = \frac{\mathbf{d}^{\dagger}\mathbf{z}}{\mathbf{d}^{\dagger}\mathbf{d}} \quad . \quad (A.9)$$

Substituting Equation (A.9) into Equation (A.7) yields

$$\Lambda = \frac{\mathbf{z}^{\dagger}\mathbf{z} + \sum_{i=1}^{K} \mathbf{z}_i^{\dagger}\mathbf{z}_i}{\mathbf{z}^{\dagger}\mathbf{z} - \frac{|\mathbf{d}^{\dagger}\mathbf{z}|^2}{\mathbf{s}^{\dagger}\mathbf{d}} + \sum_{i=1}^{K} \mathbf{z}_i^{\dagger}\mathbf{z}_i} \mathop{\gtrless}_{H_0}^{H_1} \alpha \quad . \quad (A.10)$$

Or rewriting yields

$$\frac{|\mathbf{d}^{\dagger}\mathbf{z}|^2}{\mathbf{d}^{\dagger}\mathbf{d}} \mathop{\gtrless}_{H_0}^{H_1} \frac{\alpha - 1}{\alpha}[\mathbf{z}^{\dagger}\mathbf{z} + \sum_{i=1}^{K} \mathbf{z}_i^{\dagger}\mathbf{z}_i] \quad . \quad (A.11)$$

Thus, we are comparing the power of the estimate of b with the sum of the power of the other vector components, which we will see is a simple scalar CFAR detector.

The vector notation for the decision rule can now be rewritten where the test will be more readily recognized as a scalar CFAR test, and the performance may be found.

Perform a unitary transform on the data of the form

$$\mathbf{y} = \mathbf{U}^\dagger \mathbf{z} \quad \mathbf{U} = [\mathbf{d} \vdots \mathbf{U}_1] \qquad , \qquad (A.12)$$

and $\mathbf{Y} = \mathbf{U}^\dagger \mathbf{Z}$. The normalizing assumption will be made that $\mathbf{d}^\dagger \mathbf{d} = 1$. The test can be rewritten in this notation as

$$|\mathbf{d}^\dagger \mathbf{U} \mathbf{U}^\dagger \mathbf{z}|^2 = |\mathbf{d}^\dagger \mathbf{U} \mathbf{y}|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\alpha - 1}{\alpha}[\mathbf{y}^\dagger \mathbf{y} + \sum_{i=1}^{K} \mathbf{y}_i^\dagger \mathbf{y}_i] \qquad . \qquad (A.13)$$

$\mathbf{d}^\dagger \mathbf{U}$ is the elementary vector $[1, 0, ..., 0]$, so the test can be written in terms of the vector elements as

$$|y_1|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\alpha - 1}{\alpha}[\sum_{j=1}^{N} |y_j|^2 + \sum_{i=1}^{K} \sum_{j=1}^{N} |Y_{ij}|^2] \qquad , \qquad (A.14)$$

or

$$|y_1|^2 \underset{H_0}{\overset{H_1}{\gtrless}} (\alpha - 1)[\sum_{j=2}^{N} |y_j|^2 + \sum_{i=1}^{K} \sum_{j=1}^{N} |Y_{ij}|^2] \qquad . \qquad (A.15)$$

This form of the detector is now easily recognized as a scalar CFAR detector. The data can be normalized so that the variance of the y_j's is unity. Then the y_j's are then distributed $N(0, 1)$, and y_1 is distributed $N(0, 1)$ or $N(\frac{b}{\sigma}, 1)$.

Under H_0, $|y_1|^2$ is distributed chi-square with 2 degrees of freedom, $\sum_{j=2}^{N} |y_j|^2$ is distributed chi-square with $2(N-1)$ degrees of freedom, and $\sum_{i=1}^{K} \sum_{j=1}^{N} |Y_{ij}|^2$ distributed chi-square with $2KN$ degrees of freedom. The right-hand side of Equation (A.15) has $2[(K+1)N - 1]$ degrees of freedom, and we may use these distributions to evaluate the performance. The probability of false alarm is

123

$$PFA_{sigI} = P(|\mathbf{y}_1|^2 > (\alpha - 1)T) = \int_0^\infty \int_{(\alpha-1)}^\infty \frac{T^{L-1}}{L!} e^{-T} e^{-w} dw dT = \frac{1}{\alpha^L} \quad , \qquad \text{(A.16)}$$

where $L = (K + 1)N - 1$

The probability of detection is found through comparison with Kelly [32]:

$$PD_{sigI} = 1 - \frac{1}{\alpha^L} \sum_{k=1}^{L} \binom{L}{k} (\alpha - 1)^k G_k\left(\frac{a}{\alpha}\right), \quad G_k(y) = e^{-y} \sum_{n=0}^{k-1} \frac{y^n}{n!} \quad , \quad \text{(A.17)}$$

where a is the signal-to-noise ratio $|b|^2/\sigma^2$.

If the covariance estimate based on this model is used to perform adaptive beamforming, then there will be no loss in signal-to-noise ratio. Looking at the loss factor density,

$$\rho = \frac{|\mathbf{d}^\dagger \hat{\mathbf{R}}^{-1} \mathbf{d}|^2}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d} \mathbf{d}^\dagger \hat{\mathbf{R}}^{-1} \mathbf{R} \hat{\mathbf{R}}^{-1} \mathbf{d}} \quad , \qquad \text{(A.18)}$$

and substituting $\sigma^2 \mathbf{I}$ for $\hat{\mathbf{R}}$, the result is

$$\rho = \frac{|\mathbf{d}^\dagger \mathbf{d}|^2}{\mathbf{d}^\dagger \mathbf{R}^{-1} \mathbf{d} \mathbf{d}^\dagger \mathbf{R} \mathbf{d}} \quad . \qquad \text{(A.19)}$$

This is independent of the estimate of σ^2 and is unity provided that the model is matched. When the unconstrained maximum-likelihood estimate of the covariance matrix was used to perform beamforming the loss factor is a random variable [10]. For this case, the loss factor is not a random variable.

GLOSSARY

N	Number of sensors, data vector length
G	Number of non-zero mean data vectors
K	Number of zero mean data vectors
P	Number of data vectors $(K + G)$
M	Number of terms in finite sum approximation
L	Dimension of covariance matrix vector space
r	Sample index
ω_o	Carrier frequency
θ, φ	Physical angles relative to array reference
$S(\omega, \theta, \phi)$	Spatial-temporal power-spectral density
γ	Threshold
Λ	Likelihood ratio
l	Log-likelihood
$\mathbf{z}, (\mathbf{z}_n)$	Sampled data vector (at time n)
$\mathbf{Z}$	Data matrix containing column vectors $\mathbf{z}_1 \ldots \mathbf{z}_P$
$\mathbf{B}$	Matrix which is $E(\mathbf{Z})$
b	Scalar portion of a structured mean, $E(\mathbf{z}) = b\mathbf{d}$
$\mathbf{R}$	Covariance matrix
$\hat{\mathbf{R}}$	Estimated covariance matrix
$\mathbf{s}$	Array coordinate vector
$\mathbf{d}, \mathbf{p}, \mathbf{q}$	Array steering vector
$\mathbf{D}$	Matrix whose columns are array steering vectors
$\mathbf{\Sigma}$	Diagonal matrix of discrete spectral weights
$\mathbf{U}$	Transformation, usually a unitary matrix
Ω_s	The space of possible covariance matrices for a particular array geometry
a	Signal-to-noise ratio available using optimal beamformer

GLOSSARY
(Continued)

ρ	Ratio of signal-to-noise ratio for a linear beamformer to optimal signal-to-noise ratio
$\Re\{\}$	Real value of argument
$\Im\{\}$	Imaginary value of argument

REFERENCES

1. S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, New Jersey: Prentice-Hall Inc. (1986).

2. H.L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, New York, New York: John Wiley and Sons, Inc. (1968).

3. M.L. Skolnik, *Introduction to Radar Systems*, St. Louis. Missouri: McGraw-Hill, Inc. (1980).

4. H.L. Van Trees, *Detection. Estimation, and Modulation Theory, Part III*, New York, New York: John Wiley and Sons, Inc. (1971).

5. N.R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution," *Annals of Math. Stat.* 34, 152–177 (1963).

6. J.I. Marcum, "A statistical theory of target detection by pulsed radar." *IRE Trans. Inform. Theory* 6, 65–83 (1960).

7. B.D. Van Veen and K.M. Buckley, Beamforming: "A versatile approach to spatial filtering." *IEEE ASSP Mag.* pp. 4–24 (April 1988).

8. T. Kailath, *Lectures on Wiener and Kalman Filtering*, New York, New York: Springer-Verlag Inc. (1981).

9. N. Levinson, "The Wiener RMS (root mean square) error criterion in filter design and prediction." *J. Math. Phys.* 25, 261–278 (1947).

10. I.S. Reed, S.D. Mallett, and L.E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.* AES-10, 853–863 (1974).

11. J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE* 5, 1408–1418 (1969).

12. S.M. Kay, *Modern Spectral Estimation. Theory and Application*, Englewood Cliffs, New Jersey: Prentice-Hall Inc. (1988).

13. T.W. Anderson, *An Introduction to Multivariate Statistics*, New York, New York: John Wiley and Sons, Inc. (1984).

14. E.J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.* AES-22, 115 (1986).

15. E.J. Kelly, "Adaptive detection in non-stationary interference, part I and part II," MIT Lincoln Laboratory, Lexington, Mass.. Technical Rep. 724, (25 June 1985). DTIC AD-158810.

16. F.C. Robey, D.R. Fuhrmann, E.J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector." *IEEE Trans. Aerosp. Electron. Syst.*, to be published.

17. D.R. Fuhrmann (private communication, 1988).

REFERENCES
(Continued)

18. D.M. Boroson. "Sample size considerations for adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.* AES-16, 446–451 (1980).

19. W.F. Gabriel. "Spectral analysis and adaptive array superresolution techniques," *Proc. IEEE* 68, 654-666 (1980).

20. W.F. Gabriel. "Using spectral estimation techniques in adaptive processing antenna systems," *IEEE Trans. Antennas Propag.* AP-34, 291–300 (1986).

21. E.J. Kelly. "Performance of an adaptive detection algorithm: Rejection of unwanted signals," *IEEE Trans. Aerosp. Electron. Syst.* 25, 122-133 (1989).

22. G. Bienvenu and K. Laurent, "Optimality of high resolution array processing using the eigensystem approach," *IEEE Trans. Acoust. Speech Signal Process.* ASSP-31. 1235-1248 (1983).

23. E.J. Kelly and A.O. Steinhardt (private communication. 1989).

24. Y. Bresler, "Maximum likelihood estimation of a linearly structured covariance with application to antenna array processing," In *IEEE ASSP Workshop on Spectrum Estimation and Modeling*. Minneapolis, Minnesota (1988), pp. 172-175.

25. D.B. Rubin and T.H. Szatrowski, "Finding maximum likelihood estimates of patterned covariance matrices by the EM algorithm," *Biometrika* 69. 657–660 (1982).

26. M.I. Miller and D.L. Snyder, "The role of likelihood and entropy in incomplete-data problems applications to estimating point-process intensities and Toeplitz constrained covariances." *Proc. IEEE* 75, 892-907 (1987).

27. D.B. Williams and D.H. Johnson, "Robust maximum likelihood estimation of structured covariance matrices," In *Proc. ICASSP* (1988).

28. J. Burg, D. Luenberger, and D. Wenger, "Estimation of structured covariance matrices." *Proc. IEEE* 70, 963–974 (1982).

29. D.R. Fuhrmann, "Application of Toeplitz covariance estimation to adaptive beamforming and detection," *IEEE Trans. Acoust. Speech Signal Process.*, to be published.

30. E.J. Kelly. "Adaptive detection in non-stationary interference, part III," MIT Lincoln Laboratory, Lexington, Mass., Technical Rep. 761. (24 August 1987).

31. M. Abramowitz and I.A. Stegun, "Handbook of Mathematical Functions," National Bureau of Standards, page 505 (1970).

32. E.J Kelly, "Finite sum expressions for signal detection probabilities." MIT Lincoln Laboratory, Lexington, Mass., Technical Rep. 566 (20 May 1981). DTIC AD-A102143/5.

REFERENCES
(Continued)

33. D.A. Shnidman, "Efficient evaluation of the probability of detection and the generalized Q-function," *IEEE Trans. Inf. Theory* IT-22, 746–751 (1976).

34. E.J. Kelly (private communication, 1989).

35. E.J. Kelly and K.M. Forsythe, "Adaptive detection and parameter estimation for multidimensional signal models," MIT Lincoln Laboratory, Lexington, Mass., Technical Rep. 848 (19 April 1989). DTIC AD-A208971.

36. G.H. Golub and C.F. Van Loan, *Matrix Computations*, second edition, Baltimore, Maryland: John Hopkins University Press (1989).

37. S.W. Lang and J.H. McClellan, "Spectral estimation for sensor arrays," *IEEE Trans. Acoust. Speech Signal Process.* ASSP-31, 349-358 (1983).

38. S.U. Pillai, *Array Signal Processing*, New York, New York: Springer-Verlag Inc. (1989).

39. D.R. Fuhrmann (private communication, 1989).

40. D.G. Luenberger, *Optimization by Vector Space Methods*, New York, New York: John Wiley and Sons, Inc. (1969).

41. A.W. Roberts and D.E. Varberg, *Convex Functions*, New York, New York: Academic (1973).

42. T. Oda, *Convex Bodies and Algebraic Geometry*, New York, New York: Springer-Verlag (1985).

43. S.D. Silvey, *Optimal Design*, London, England: Chapman and Hall (1980).

44. K. Glashoff and S. Gustafson, *Linear Optimization and Approximation*, New York, New York: Springer-Verlag Inc. (1983).

45. U. Grenander and G. Szegö, *Toeplitz Forms and their Applications*, New York, New York: Chelsea Publishing Company (1984).

46. V.F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. R. Astronom. Soc.* 33, 347–366 (1973).

47. A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statis.* B-39, 1–37 (1977).

48. P. Moulin, D.L. Snyder, and J.A. O'Sullivan, "Maximum-likelihood spectrum estimation of periodic processes from noisy data," to be published.

49. D.R. Fuhrmann and M.I. Miller, "On the existence of positive definite maximum-likelihood estimates of structured covariance matrices," *IEEE Trans. Inf. Theory* 34, 722-729 (1988).

REFERENCES
(Continued)

50. M.I. Miller and D.R. Fuhrmann, "Maximum-likelihood narrow-band direction finding and the EM algorithm," *IEEE Trans. Acoust. Speech Signal Process.* 38, 1560-1577 (1990).

51. K. Senne (private communication, 1989).

52. E.L. Lehmann, *Testing Statistical Hypothesis.* New York, New York: John Wiley and Sons, Inc. (1986).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 30 July 1991	3. REPORT TYPE AND DATES COVERED Technical Report

4. TITLE AND SUBTITLE

A Covariance Modeling Approach to
Adaptive Beamforming and Detection

5. FUNDING NUMBERS

C — F19628-90-C-0002

6. AUTHOR(S)

Frank C. Robey

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Lincoln Laboratory, MIT
P.O. Box 73
Lexington, MA 02173-9108

8. PERFORMING ORGANIZATION REPORT NUMBER

TR-918

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Air Force
Electronic Systems Division
Hanscom AFB, MA 01730

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ESD-TR-91-056

11. SUPPLEMENTARY NOTES

None

12a. DISTRIBUTION AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The subject of this report is the general problem of signal processing for sensor arrays. Under certain reasonable assumptions, the model for the noise covariance matrix of the vector of array outputs is an integral involving the spatial-temporal power-spectral-density function. This report examines the application of this covariance model to problems in adaptive beamforming and detection. A constant false alarm rate detector, based on unconstrained maximum-likelihood techniques, is derived and analyzed. The space of noise covariance matrices possible from a particular array is characterized, yielding representations for the space and members of the space in terms of finite numbers of spectral points. These representations are used to derive constrained maximum-likelihood estimators that jointly estimate the parameters of the density function. Two approaches that use the constrained covariance estimates to perform beamforming are described and compared. The loss in signal-to-noise ratio and the variance of the estimators are shown to be less for these approaches than for those that do not use the covariance model. Detection methods based on the generalized likelihood ratio test and a constant false alarm rate matched-filter detector are analyzed, and simulation results are presented.

14. SUBJECT TERMS

array signal processing, maximum likelihood, covariance estimator, adaptive beamforming, adaptive detection, statistical hypothesis testing, adaptive matched filter, estimation, detection

15. NUMBER OF PAGES

150

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	None